

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 8, 2005, 03:43:40 ; Search time 271.773 Seconds

(without alignments)
5783.472 Million cell updates/sec

Title: US-10-728-051-1

Perfect score: 3562

Sequence: 1 aataatcatatattcatc.....cgtttgctggtttttctcc 2032

Scoring table:

BLOSUM62
-Qs/cpn2 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2105692 segs, 386760381 residues

Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Qs/cpn2 1/USPTO spool_p/US10728051/runat_07042005_125044_14041/app_query.fasta_1.4757
-DB=A_Geneseq -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10728051 @CGN 1 1 672 @runat_07042005_125044_14041 -NCPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq.i6Dec04.*

- 1: Geneseq1980s.*
- 2: Geneseq1980s.*
- 3: Geneseq2000s.*
- 4: Geneseq2001s.*
- 5: Geneseq2002s.*
- 6: Geneseq2003as.*
- 7: Geneseq2003bs.*
- 8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3286	92.3	626	2 AAY15244	Aay15244 Peanut al
2	3286	92.3	626	2 AAY25657	Aay25657 Peanut al
3	3286	92.3	626	6 ABUS2412	Abus2412 Peanut Ar
4	3286	92.3	626	7 ADC34906	Adc34906 Peanut al
5	3282	92.1	626	4 AAU04706	Aau04706 Anaphylac
6	3282	92.1	626	6 ABUS2570	Abus2570 Peanut Ar
7	3280	92.1	626	6 ABUS2568	Abus2568 Peanut Ar
8	3280	92.1	626	6 ABUS2571	Abus2571 Peanut Ar
9	3280	92.1	626	6 ABUS2574	Abus2574 Peanut Ar
10	3280	92.1	626	6 ABUS2573	Abus2573 Peanut Ar

11	3279	92.1	626	8 ADM12053	Adm12053 Arachis h
12	3278	92.0	626	6 ABUS2569	Abus2569 Peanut Ar
13	3278	92.0	626	6 ABUS2572	Abus2572 Peanut Ar
14	3274	91.9	626	2 AAW22150	Aaw22150 Peanut al
15	3269.5	91.8	625	7 ADG27464	Adg27464 Peanut Ar
16	3251	91.3	626	8 ADO38271	Ado38271 Peanut al
17	3052	85.7	634	3 AAB33599	Aab33599 Modified
18	3052	85.7	634	4 AAU04709	Aau04709 Modified
19	3052	85.7	634	4 AAU05034	Aau05034 Modified
20	3041	85.4	614	2 AAW22149	Aaw22149 Peanut al
21	3041	85.4	614	2 AAW62834	Aaw62834 Arachis h
22	3041	85.4	614	2 AAW52413	Aaw52413 Peanut Ar
23	3041	85.4	614	7 ADG27465	Adg27465 Peanut Ar
24	2130	59.8	415	2 AAY40913	Aay40913 Ara h 1 a
25	1281	36.0	605	2 AAW62838	Aaw62838 Glycine m
26	1281	36.0	605	7 ADG27564	Adg27564 Soybean B
27	1276	35.8	605	2 AAY40999	Aay40999 Soybean b
28	1190	33.4	559	5 ABG71270	Abg71270 Glycine m
29	1180.5	33.1	543	5 ABG71271	Abg71271 Glycine m
30	1165	32.7	417	5 ABG71269	Abg71269 Glycine m
31	1157	32.5	417	6 ABP97239	Abp97239 Mature be
32	1157	32.5	425	6 ABP97238	Abp97238 FLAG-tag
33	1156	32.5	439	6 ABP97235	Abp97235 Glycine m
34	1117.5	31.4	390	6 ABP97236	Abp97236 Glycine m
35	865.5	24.3	666	2 AAW62829	Aaw62829 Macadamia
36	863.5	24.2	625	2 AAW62830	Aaw62830 Macadamia
37	849.5	23.8	666	2 AAW62828	Aaw62828 Macadamia
38	811	22.8	566	2 AAR20181	Aar20181 Sequence
39	791	22.2	590	2 AAW62832	Aaw62832 Gossypium
40	764	21.4	371	2 AAY40914	Aay40914 Phaseolin
41	737.5	20.7	525	2 AAW62831	Aaw62831 Theobroma
42	654	18.4	366	6 ABUS2503	Abus2503 Soybean B
43	629.5	17.7	593	2 AAW62835	Aaw62835 Zea mays
44	626	17.6	573	7 ADL18493	Adl18493 Maize glo
45	614	17.2	637	2 AAW62837	Aaw62837 Hordeum v

ALIGNMENTS

RESULT 1

AA15244
ID AAY15244 standard; protein; 626 AA.

XX AC AAY15244;
XX AC
XX DT 17-OCT-2003 (revised)
XX DT 09-NOV-1999 (first entry)
XX DE Peanut allergen, Ara h 1, amino acid sequence.
XX KW allergy; immune response; transgenic; allergen; epitope;
XX KW immunoglobulin E; Ig E; binding site; peanut.
XX OS Arachis hypogaea.
XX PN WO9838978-A1.
XX PD 05-AUG-1999.
XX PF 29-JAN-1999; 99WO-US002031.
XX PR 31-JAN-1998; 98US-0073283P.
XX PR 13-FEB-1998; 98US-0074590P.
XX PR 13-FEB-1998; 98US-0074624P.
XX PR 13-FEB-1998; 98US-0074633P.
XX PR 27-AUG-1998; 98US-00141220.
XX PA {UAY-} UNIV ARKANSAS.
XX PA {UNY } UNIV NEW YORK MT SINAI SCHOOL MEDICINE.
XX PA {SOSI/} SOSIN H.
XX PI Sosin H, Bannon GA, Burks AW, Sampson HA;
XX

DR WPI; 1999-479189/40.
DR N-PSDB; AA206382.
XX Modified allergen with reduced IgE binding, useful for treating e.g.
PT allergies.
XX
PS Disclosure; Page 35-37; 46pp; English.
XX
CC This is the amino acid sequence of the Ara h 1 protein from Arachis
CC hypogaea. The Ara h 1 protein has 23 IgE (Immunoglobulin E) binding
CC epitopes, four of which are immunodominant (AAV15247, AAV15249, AAV15250
CC and AAV15263). By modifying the IgE binding sites the ability of the
CC allergen to provoke an immune response is downregulated. The epitopes of
CC the IgE binding sites can therefore be modified in genetically engineered
CC plants and animals to elicit less of an allergic response. (Updated on 17
CC -Oct-2003 to standardise OS field)
XX
SQ Sequence 626 AA;

Alignment Scores:
Pred. No.: 1.4e-304 Length: 626
Score: 3286.00 Matches: 626
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 92.25% Indels: 0
DB: 2 Gaps: 0

US-10-728-051-1 (1-2032) x AAV15244 (1-626)

QY 50 ATGAGAGGAGGGTTCTCCACATGATGCTGTGCTAGGATCCTTGCTGCTCAGTT 109
DB 1 MetArgGlyArgValSerProLeuMetLeuLeuLeuGlyIleLeuValLeuAlaSerVal 20
QY 110 TCTGAACCCATGCAAGTCATCACCTTACCAAGAGAAACAGAGAACCCCTGCCCCAG 169
DB 21 SerAlaThrHisAlaLysSerProTyrGlnLysLysThrGluAsnProCysAlaGln 40
QY 170 AGGTGCTCCAGAGTTGTCAACAGGAACCGGATGACTTGAAGCAAAAGCATCGAGTCT 229
DB 41 ArgCysLeuGlnSerCysGlnGlnGluProAspAspLeuLysGlnLysAlaCysGluSer 60
QY 230 CGCTGCACCAAGCTCGAGTATGATCCTGTTGTGCTATGATCCTCGAGGACACACTGGC 289
DB 61 ArgCysThrLysLeuGluTyrAspProArgCysValTyrAspProArgGlyHisThrGly 80
QY 290 ACCACCAACCAAGTTCCCTCCAGGGAGCGGACACAGTGGCGGCCCAACCCGGAGACTAC 349
DB 81 ThrThrAsnGlnArgSerProProGlyGluArgThrArgGlyArgGlnProGlyAspTyr 100
QY 350 GATGATGACCGCGCTCAACCCCAAGAGAGGAGGAGCGCGATGGGACCACTGGACCG 409
DB 101 AspAspAspArgArgGlnProArgArgGluGluGlyGlyArgTyrGlyProAlaGlyPro 120
QY 410 AGGGAGCTGAAGAGAGAAAGACTGGAGACCAACCAAGAGAGATTGGAGGCGACCAAGT 469
DB 121 ArgGluArgGluArgGluAspTyrArgGlnProArgGluAspTyrArgArgProSer 140
QY 470 CATCAGACCCAGGAAATAAGGCCCAAGAGAGAGAGAGAGAAACAGAGTGGGGAACA 529
DB 141 HisGlnGlnProArgLysLysLeuArgProGluGlyArgGluGlyGluGlnGluTrpGlyThr 160
QY 530 CCAGGTAGCCATGTGAGGGAAGAAACATCTCGAAACAAACCCCTTTCTACTTCCCGTCAAGG 589
DB 161 ProGlySerHisValArgGluGluThrSerArgAsnAsnProPheTyrPheProSerArg 180
QY 590 CGGTTTAGCCCGCTACGGGACCAACCAACCGTAGGATCCGGTCTCGAGAGTTGAC 649
DB 181 ArgPheSerThrArgTyrGlyAsnGlnAsnGlyArgIleArgValLeuGlnArgPheAsp 200
QY 650 CAAAGTCAAGGCAGTTTCCAGATCTCCAGAAATCACCGTATTGTGAGATCCAGGCCAAA 709
DB 201 GlnArgSerArgGlnPheGlnAsnLeuGlnAsnHisArgIleValGlnIleGluAlaLys 220

QY 710 CTTAACACTCTGTGTTCTTCCCAAGCACGCTGATGCTGATAACATCTTGTATTATCAGCAA 769
DB 221 ProAsnThrLeuValLeuProLysHisAlaAspAlaAspAsnIleLeuValIleGlnGln 240
QY 770 GGGCAGCCACCGTACCGTAGCAATAAGCAATACAGAAAGAGCTTTAACTTTCACGAG 829
DB 241 GlyGlnAlaThrValThrValAlaAsnGlyAsnAsnArgLysSerPheAsnLeuAspGlu 260
QY 830 GGCCATGCACCTCAGAAATCCCATCCGTTTCTATCTTCTACATCTTGAACCGCATCACAAC 889
DB 261 GlyHisAlaLeuArgIleProSerGlyPheIleSerTyrIleLeuAsnArgHisAspAsn 280
QY 890 CAGAACCTCAGAGTAGCTAAATCTTCATGCCGCTTAAACACACCCGCCAGCTTTCAGAT 949
DB 281 GlnAsnLeuArgValAlaLysIleSerMetProValAsnThrProGlyGlnPheGluAsp 300
QY 950 TTCTTCCCGGAGCAGCGGAGACCAATCATCTTCTGAGGCTTTCAGCAGGATACG 1009
DB 301 PhePheProAlaSerSerArgAspGlnSerTyrLeuGlnGlyPheSerArgAsnThr 320
QY 1010 TTGGAGCGCGCTTCAATGCCGAATTCATCAGATACCGAGGCTGCTGTAGAACAGAAAT 1069
DB 321 LeuGluAlaAlaPheAsnAlaGluPheAsnGluIleArgValLeuLeuGluGluAsn 340
QY 1070 GCAGAGGTGAGCAAGAGAGAGGAGGCGAGGATGAGTACTCGAGTAGTGAGAAC 1129
DB 341 AlaGlyGlyGluGlnGluGluArgGlyGlnArgTyrSerThrArgSerSerGluAsn 360
QY 1130 AATGAGAGGTAGTAGTCAAAAGTGTCAAGAGAGACGCTTGAAGACTTACTTAAGCACGCT 1189
DB 361 AsnGluGlyValIleValLysValSerLysGluHisValGluGluLeuThrLysHisAla 380
QY 1190 AAATCCGCTCTCAAGAAAGGCTCCGAAAGAGGAGAGATATCACCAACCCCAATCAACTTG 1249
DB 381 LysSerValSerLysLysGlySerGluGluGluGlyAspIleThrAsnProIleAsnLeu 400
QY 1250 AGAGAGCGGAGCGCGATCTTTTAAACAACTTGGAAAGTTATTGAGGTGAGCGACAC 1309
DB 401 ArgGluGlyGluProAspLeuSerAsnAsnLysLysLeuPheGluValLysProAsp 420
QY 1310 AAGAGAACCCCGCTCAGAGCTCAGACATGATGCTCACCTGTGTAGAGATCAAGAA 1369
DB 421 LysLysAsnProGlnLeuGlnAspLeuAspMetLeuThrCysValGluLysGlu 440
QY 1370 GGAGCTTTGATGCTCCCACTTCAACTCAAGGCCATGTTATCGTCTCGTCAACAAA 1429
DB 441 GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValAsnLys 460
QY 1430 GGAATCGAAACCTTGAATCTGCTGCTGTAGAAAGAGCAACACAGAGGCGGCGCG 1489
DB 461 GlyThrGlyAsnLeuGluValAlaValArgLysGluGlnGlnArgGlyArg 480
QY 1490 GAAGAAGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1549
DB 481 GluGluGluGluAspGluAspGluGluGluGluGluGluGluGluGluValArgArgTyr 500
QY 1550 ACAGCAGGTTTGAAGGAAGGCGATGTTTCTATCGCCAGCAGCTCATCCAGTAGCCATC 1609
DB 501 ThrAlaArgLeuLysGluYAspValPheIleMetProAlaAlaHisProValAlaIle 520
QY 1610 AACGTTCTCGAACTCCATCTGCTGCTTCCGTTATCAACGCTGAACCAACACACAGA 1669
DB 521 AsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsnAsnHisArg 540
QY 1670 ATCTTCTTTCAGGTTGAAGCAATGTGTATAGACACAGATAGAGAGAGAGAGAGAG 1729
DB 541 IlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGlnAlaLysAsp 560
QY 1730 TTAGCATTTCCCTGGGTGAAACAGTTGAGAGCTCATCAAAACACAGAGGAATCT 1789
DB 561 LeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGlnLysGluSer 580
QY 1790 CACTTTGTGAGTGTGCTGCTCCTCAATCTCAATCTCCGTCCTCTCTCTGAGAAAGAG 1849

Qy	290	ACCAACCAACCAACCTTCCCTCCAGGGAGCGGACACGTGGCCGCCCAACCCGAGACTAC	349
Db	81	ThrThrAsnGlnArgSerProProGlyGluArgThrArgGlyArgGlnProGlyAspTyr	100
Qy	350	GATGATCAGCCGCGTCAACCCGAGAGAGAGAGGAGCGCGATGGGACACAGCTGGACCG	409
Db	101	AspAspAspArgArgGlnProArgArgGluGluGlyArgTrpGlyProAlaGlyPro	120
Qy	410	AGGAGCGGTGAAGAAGAAGAGACTGGAGAGCAACCAAGAGAGAATTCGAGCGCCACCAAGT	469
Db	121	ArgGluArgGluArgGluAspTrpArgGlnProArgGluAspTrpArgArgProSer	140
Qy	470	CATCAGCAGCCACGGAAAAATAAGCCCGAAGGAAGAGAGAGAACAAAGAGTGGGGAAACA	529
Db	141	HisGlnGlnProArgLysIleArgProGluGlyArgGluGlyGluGlnGluTrpGlyThr	160
Qy	530	CCAGGTAGCCATGTAGGGAGAGAACATCTCGGAACAACCTTCTACTTCCCGTCAAGG	589
Db	161	ProGlySerHisValArgGluGluThrSerArgAsnAsnProPheTyrPheProSerArg	180
Qy	590	CGGTTTAGCACCCCTCAACGGGAAACCAAAACGGTAGGATCCGGGTCTCGACAGGTTTGAC	649
Db	181	ArgPheSerThrArgTyrGlyAsnGlnAsnGlyArgIleArgValLeuGlnArgPheAsp	200
Qy	650	CAAGGTCAGGCGAGTTTCAGAAATCTCCAGAATCACCGTATGTGCAGATCGAGGCCAAA	709
Db	201	GlnArgSerArgGlnPheGlnAsnLeuGlnAsnHisArgIleValGlnIleGluAlaLys	220
Qy	710	CCTTAACACTCTGTCTTCCCAAGCAGCTGATGCTGATACATCTCTTGTTATCCAGCAA	769
Db	221	ProAsnThrLeuValLeuProLysHisAlaAspAlaAspAsnIleLeuValIleGlnGln	240
Qy	770	GGGCAAGCCACCGTCAGCGTAGCAAAATGGCAATAACAGAAAGAGCTTTAATCTTGACGAG	829
Db	241	GlyGlnAlaThrValThrValAlaAsnGlyAsnAsnArgLysSerPheAsnLeuAspGlu	260
Qy	830	GGCCATGCACCTCAGAAATCCCATCCGTTTTCATTTCTACATCTTTGAACCCGCCATGCAAC	889
Db	261	GlyHisAlaLeuArgIleProSerGlyPheIleSerTyrIleLeuAsnArgHisAspAsn	280
Qy	890	CAGAACCTCAGAGTAGCTAAATCTCCATCCGCGTTAAACACACCCGCCAGTTTGAGGAT	949
Db	281	GlnAsnLeuArgValAlaLysIleSerMetProValAsnThrProGlyGlnPheGluAsp	300
Qy	950	TTCTTCCGGCGAGCAGCCGAGACCAATCATCTTCTTACCTTGCAGGCGTTTCAGCAGGAATACG	1009
Db	301	PhePheProAlaSerSerArgAspGlnSerSerTyrLeuGlnGlyPheSerArgAsnThr	320
Qy	1010	TTGAGGCGCGCTTCAATGCGGAATTCATATGAGATACGAGGCGTCTGTTAGAGAGAAAT	1069
Db	321	LeuGluAlaAlaPheAsnAlaGluPheAsnGluIleArgArgValLeuLeuGluAsn	340
Qy	1070	GCAGGAGGTGACGAAGAGGAGAGAGGCGCAGAGCGATGGAGTACTCGGAGTAGTGAGAAC	1129
Db	341	AlaGlyGlyGluGlnGluGluArgGlyGlnArgArgTrpSerThrArgSerSerGluAsn	360
Qy	1130	AATCAAGGAGTGATAGTCAAGTGTCAAGAGGAGCAGCTTCAAGAACCTTACTAAGCAGCGCT	1189
Db	361	AsnGluGlyValIleValLysValSerLysGluHisValGluGluLeuThrLysHisAla	380
Qy	1190	AAATCCCGTCTCAAGAAAGGCTCCGAAGAAGAGGGAGATATCACCAACCCAAATCAACTTG	1249
Db	381	LysSerValSerLysLysGlySerGluGluGluGlyAspIleThrAsnProIleAsnLeu	400
Qy	1250	AGAGAAGCGAGCCCGATCTTTCTTAAACAATTTGGGAAGTTATTTAGGTGAAGCCAGAC	1309
Db	401	ArgGluGlyGluProAspLeuSerSerAsnAsnPheGlyLysLeuPheGluValLysProAsp	420
Qy	1310	AAGAAGAACCCCGACTTCAGGACCTGGACATGATGCTCACCTGTGTAGAGATCAAGAA	1369
Db	421	LysLysAsnProGlnLeuGlnAspLeuAspMetMetLeuThrCysValGluIleLysGlu	440

Qy	1370	GGAGCTTTGATGTCCTCCACACTTCAACTAAAGGCCCATGGTTATTCGTCGTCGTCACAAA	1429
Db	441	GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValAsnLys	460
Qy	1430	GGAACTGGAAACCTTGAACTCGTGGCTGTAAGAAAAGAGCAACAACAGAGGGGACGCGG	1489
Db	461	GlyThrIlyAsnLeuGluLeuValAlaValArgLysGluGlnGlnArgIlyArg	480
Qy	1490	GAAGAAGAGGAGGACGAAGACGAAGAAGAGAGGAGGGAAGTAAACAGAGAGGTGCGTAGGTAC	1549
Db	481	GluGluGluGluAspGluAspGluGluGluGluGlySerAsnArgGluValAlaArgGlyTyr	500
Qy	1550	ACAGCAGGTTGAAGAGAGCGGATGTGTTTCATCATGCCAGCAGCTATCCAGTAGCCATC	1609
Db	501	ThrAlaArgLeuLysGluGlyAspValPheIleMetProAlaAlaHisProValAlaIle	520
Qy	1610	AAGCTTCTCCGAACCTCCATCTGCTTGGCTTCGGTATCAACGCTGAAAAACAACACACAGA	1669
Db	521	AsnAlaSerSerGluLeuHisLeuLeuGlyPheIlyIleAsnAlaGluAsnAsnHisArg	540
Qy	1670	ATCTTCCTTCGACGTCATTAAGGACAATGTGATAGACCAGATAGAGAAAGCAAGCGAAGGAT	1729
Db	541	IlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGlnAlaLysAsp	560
Qy	1730	TTAGCATTCCTCGGTGCGGTGAACAAGTTGAGAGAGCTCATCAAAAACAGAGAGGAATCT	1789
Db	561	LeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGlnLysGluSer	580
Qy	1790	CACCTTGTGAGTGCTCGTCCTCAATCTCAATCTCAATCTCGTCGTCCTCTGAGAAAGAG	1849
Db	581	HisPheValSerAlaArgProGlnSerGlnSerGlnSerProSerSerProGluLysGlu	600
Qy	1850	TCTCCTCAGAAAAGAGGATCAAGAGGAGGAAAAACAAGGAGGAGGTCCTCCTCTGAGAAAG	1909
Db	601	SerProGluLysGluAspGlnGluGluGluAsnGlnGlyLysGlyProLeuLeuSer	620
Qy	1910	ATTTTGAAGCTTTTAAAC	1927
Db	621	IleLeuLysAlaPheAsn	626
RESULT 4			
ADC34906			
ID	ADC34906 standard; protein; 626 AA.		
XX	ADC34906;		
AC	ADC34906;		
XX	ADC34906;		
DT	18-DEC-2003 (first entry)		
XX	Peanut allergen Ara h 1.		
DE	allergen; antigen; hyporesponsive; desensitization; immunomodulator;		
KW	gene therapy; peanut.		
KW	Arachis hypogaea.		
OS	WO2003047618-A2.		
XX	WO2003047618-A2.		
XX	12-JUN-2003.		
XX	05-DEC-2002; 2002WO-GB0005548.		
PF	05-DEC-2001; 2001US-0338385P.		
XX	(CIRC-) CIRCASSIA LTD.		
PA	Larche M, Ledger PW;		
XX	WPI; 2003-523267/49.		
XX	Desensitizing an individual to a selected polypeptide antigen comprises		
PT	administering a composition containing polypeptide antigens in an amount		
PT	that generates a state of hyporesponsiveness to the antigen to allow		
PT	desensitization.		

XX PS Disclosure; Page 40; 57pp; English.

XX The invention relates to a novel method for desensitising an individual to a selected polypeptide antigen. The method comprises administering a composition that contains polypeptide antigens in an amount that generates in the individual a state of hyporesponsiveness to the antigen to allow desensitisation to one or more polypeptide antigens. The method of the invention has immunomodulator activity, and may have a use in gene therapy. The composition and method are useful in manufacturing a medicament for desensitising an individual to a selected polypeptide antigen or for generating in the individual a state of hyporesponsiveness to the antigen to allow desensitisation to one or more polypeptide antigens. The present sequence is used in the exemplification of the invention.

XX SQ Sequence 626 AA;

Alignment Scores:

Pred. No.:	1-4e-304	Length:	626
Score:	3286.00	Matches:	626
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	92.25%	Indels:	0
DB:	7	Gaps:	0

US-10-728-051-1 (1-2032) x ADC34906 (1-626)

QY 50 ATGAGAGGAGGGTTTCTCCACGTAGTCTGTTCTAGGAGATCCTGTCCTGGCTTCAGTT 109

DB 1 MetArgGlyArgValSerProLeuMetLeuLeuGlyIleLeuValLeuAlaSerVal 20

QY 110 TCTGCAAGCATGCCAAGTTCATACCTTACAGAGAAACACAGAACCCCTCGCCCGAG 169

DB 21 SerAlaThrHisAlaLysSerSerProTyrGlnLysLysThrGluAsnProCysAlaGln 40

QY 170 AGTGCCCTCCAGAGTTGTCAACAGGAACCGGATGACTTGAAGCAAAAGGCATCGAGTCT 229

DB 41 ArgCysLeuGlnSerCysGlnGlnProAspAspLeuLysGlnLysAlaCysGluSer 60

QY 230 CGTGCACCAAGCTCGAGTATGATCCTGTTGTCTATGATCCTCGAGCACACACTGGC 289

DB 61 ArgCysThrLysLeuGluTyrAspProArgCysValTyrAspProArgGlyHisThrGly 80

QY 290 ACCACCAACCAACCTTCCCTCCAGGGGAGCGGACACAGTGGCCGCCCAACCCGAGACTAC 349

DB 81 ThrThrAsnGlnArgSerProProGlyGluArgThrArgGlyArgGlnProGlyAspTyr 100

QY 350 GATGATACCCCGTCAACCCCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 409

DB 101 AspAspAspArgGlnProArgGlnProArgGlnProArgGlnProArgGlnProArgGln 120

QY 410 AGGAGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 469

DB 121 ArgGluArgGluArgGluArgGluArgGluArgGluArgGluArgGluArgGluArgGlu 140

QY 470 CATCAGCAGCCGCAAGAAATAGCCCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 529

DB 141 HisGlnGlnProArgLysLysArgProGluArgGluArgGluArgGluArgGluArgGlu 160

QY 530 CCAGTAGCCCATGTGAGGAGAGAAACATCTCGGAACACCCCTTCTACTTCCCGTCAAGG 589

DB 161 ProGlySerHisValArgGluGluThrSerArgAsnAsnProPheTyrPheProSerArg 180

QY 590 CGGTTTAGCCCGCTACCGGAACCAAAACCGTAGGATCCGGGTCCTGCAGAGTTTGAC 649

DB 181 ArgPheSerThrArgTyrGlyAsnGlnAsnGlyArgIleArgValLeuGlnArgPheAsp 200

QY 650 CAAAGGTCAAGCAGTTTCAGAAATCTCCAGAAATCAGCTATTTGTCAGATCGAGCCAAA 709

DB 201 GlnArgSerArgGlnPheGlnAsnLeuGlnAsnHisArgIleValGlnIleGluAlaLys 220

QY 710 CCTAACACTCTGTTCTTCCCAAGCAGCGCTGATGATGATGATGATGATGATGATGATGAT 769

DB 221 ProAsnThrLeuValLeuProLysHisAlaAspAlaAspAsnIleLeuValIleGlnGln 240

QY 770 GGGCAAGCCACCGTAGCCGTAATGGCAATTAACAGAAAGAGCTTTAATCTTTGACGAG 829

DB 241 GlyGlnAlaThrValThrValAlaAsnGlyAsnAsnArgLysSerPheAsnLeuAspGlu 260

QY 830 GGCCATGCACCTCAGAAATCCCATCCGTTTCAATTTCTTACATCTTTGAACCCGCTGACAAC 889

DB 261 GlyHisAlaLeuArgIleProSerGlyPheIleSerTyrIleLeuAsnArgHisAspAsn 280

QY 890 CAGAACCTCAGAGTAGTAATAATCTCCATGCCCTTAAACACACCCGCCAGTTGAGAT 949

DB 281 GlnAsnLeuArgValAlaLysIleSerMetProValAsnThrProGlyGlnPheGluAsp 300

QY 950 TTCTTCCCGCGAGCAGCCGAGACCAATCATCTTCTTACCTTTCAGGCGCTTCAGCAGGAATACG 1009

DB 301 PhePheProAlaSerSerArgAspGlnSerSerTyrLeuGlnGlyPheSerArgAsnThr 320

QY 1010 TTGGAGCGCGCTTCAATGCGGAATTCATGAGATACGGAGGGTCTGTTTGAAGAGNAT 1069

DB 321 LeuGluAlaAlaPheAsnAlaGluPheAsnGluIleArgArgValLeuLeuGluGluAsn 340

QY 1070 CCAGGAGTGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1129

DB 341 AlaGlyGlyGluGlnGluGluArgGlyGlnArgArgTrpSerThrArgSerSerGluAsn 360

QY 1130 AATGAAGAGTAGTAGTCAAAAGTGTCAAAGAGCAGCGTTGAAGAACTTCTTAAAGCACGCT 1189

DB 361 AsnGluGlyValIleValLysValSerLysGluHisValGluGluLeuThrLysHisAla 380

QY 1190 AATCCGTCTCAAAGAAAGGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1249

DB 381 LysSerValSerLysLysGlySerGluGluGluGlyAspIleThrAsnProLleAsnLeu 400

QY 1250 AGAGAGGCGAGCCCGATCTTCTAACTTGGAAAGTATTATTTAGGTGAAGCCAGAC 1309

DB 401 ArgGluGlyGluProAspLeuSerAsnAsnPheGlyLysLeuPheGluValLysProAsp 420

QY 1310 AAGAAGAACCCCGCTCAGGACTTGCAGATGATGCTCCTGCTGCTGCTGCTGCTGCTGCTGCT 1369

DB 421 LysLysAsnProGlnLeuGlnAspLeuAspMetMetLeuThrCysValGluLysGlu 440

QY 1370 GGAGCTTTCATCTCCACACTTCACTCAAGGCGATGTTTATCTGCTGCTGCTGCTGCTGCTGCT 1429

DB 441 GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValValAsnLys 460

QY 1430 GGAACCTGGAAACCTTGAACCTCGTGGCTGTAAAGAAAGAGCAACACAGAGGGGAGCGCGG 1489

DB 461 GlyThrGlyAsnLeuGluLeuValAlaValA-gLysGluGlnGlnArgGlyArgArg 480

QY 1490 GAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1549

DB 481 GluGluGluGluAspGluAspGluGluGluGluGlySerAsnArgGluValArgArgTyr 500

QY 1550 ACAGCCAGGTGTGAAGAGAGCGATGTTTCATCATGCGCCAGCAGCTCATCCAGTAGCCATC 1609

DB 501 ThrAlaArgLysGluGlyAspValPheIleMetProAlaAlaHisProValAlaIle 520

QY 1610 AACGCTTCTCCGAATCCATCTGCTTGGCTTCGGTATCAACGCTGAAACCAACACAGAG 1669

DB 521 AsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsnAsnHisArg 540

QY 1670 ATCTTCTTCAGGTGATAGGCAATGTGATAGCAGATAGAGAGAGAGAGAGAGAGAGAGAGAT 1729

DB 541 IlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGlnAlaLysAsp 560

QY 1730 TTAGCATTTCTCCGCTCGGTCGAGCAAGTTTGAGAGCTCATCAAAACCCAGAGAGATCT 1789

DB 561 LeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGlnLysGluSer 580

QY 1790 CACTTGTGAGTCTGCTCTCAATCTCAATCTCAATCTCCGTCGCTCTCTGAGAAAGAG 1849

Db 581 HisPheValSerAlaArgProGlnSerGlnSerGlnSerProSerProGluLysGlu 600
QY 1850 TCTCTCTGAGAAAGAGATCAAGAGGAGGAAACCAAGAGGAGGATCCACTCTTTCA 1909
Db 601 SerProGluLysGluAspGlnGluGluGluAsnGlnGlyGlyGlyGlyProLeuLeuSer 620
QY 1910 ATTTTGAAGGCTTTTAAAC 1927
Db 621 IleLeuLysAlaPheAsn 626

RESULT 5
AAU04706
ID AAU04706 standard; protein; 626 AA.
XX AC AAU04706;
XX DT 23-OCT-2001 (first entry)
XX DE Anaphylactic antigen Ara h 1.
XX KW Ara h 1; anaphylactic antigen; immunoglobulin E; IgE; immunogenic;
XX KW allergy; mast cell; basophil; mouse.
XX OS Mus sp.

XX PH Key Location/Qualifiers
XX FT Misc-difference 285 /note= "Encoded by gta"
XX FT WO200140264-A2.
XX PN 07-JUN-2001.
XX PF 06-DEC-2000; 2000MO-US033124.
XX PR 06-DEC-1999; 99US-00455294.
XX PR 23-JUN-2000; 2000US-0213765P.
XX PR 27-SEP-2000; 2000US-0235797P.
XX PA (PANA-) PANACEA PHARM LLC.
XX PA (UYAR-) UNIV ARKANSAS.
XX PA (MOUN) MOUNT SINAI SCHOOL MEDICINE NEW YORK.
XX PI Bannon GA, Burks WA, Caplan MJ, Sampson H, Sosin H;
XX PS WPI; 2001-381378/40.
XX DR N-PSDB; AAS08537.
XX PT Antigenic fragments useful for reducing anaphylactic risk and reducing
XX PT the severity and/or number of allergic symptoms in individuals sensitive
XX PT to antigens, have reduced ability to bind Immunoglobulin E.
XX PS Claim 7; Fig 9; 100pp; English.

XX CC The sequence represents the amino acid sequence of anaphylactic antigen
XX CC Ara h 1. Ara h 1 is an anaphylactic antigen (A), which was used to design
XX CC antigenic peptides having a reduced ability to bind immunoglobulin E
XX CC (Ige) as compared with the intact (A), or having a sequence substantially
XX CC identical to a portion of sequence of an antigen that includes at least
XX CC one Ige binding site, where at least one Ige binding site of the peptide
XX CC is altered. The antigenic peptides are used in a composition which is
XX CC useful for reducing risk or severity of allergic reaction to an antigen.
XX CC This is done by identifying an individual at risk of allergic reaction to
XX CC an antigen by identifying prior display of allergic symptoms when exposed
XX CC to the antigen, or a familial relationship with an individual who
XX CC previously displayed allergic symptoms when exposed to the antigen.
XX CC Following this an antigen-specific Ige present on one or more mast cells
XX CC or basophils in the individual's serum is identified. The individual is
XX CC then contacted with a peptide corresponding to a portion of the antigen,
XX CC which is selected, formulated, and delivered so that binding of the
XX CC peptide to antigen-specific Ige is reduced as compared with Ige binding
XX CC of intact antigen. The composition is also useful for treating and
XX CC preventing allergic reactions

XX SQ Sequence 626 AA;
Alignment Scores:
Pred. No.: 3,38e-304 Length: 626
Score: 3282.00 Matches: 625
Percent Similarity: 99.84% Conservative: 0
Best Local Similarity: 99.84% Mismatches: 1
Query Match: 92.14% Indels: 0
DB: 4 Gaps: 0

US-10-728-051-1 (1-2032) x AAU04706 (1-626)

QY 50 ATGAGAGGAGGGTTTCTCCTGATGCTTCTAGGATCCTTGTCTGGTTCAGTT 109
Db 1 MetArgGlyArgValSerProLeuMetLeuLeuGlyIleLeuValLeuAlaSerVal 20
QY 110 TCTGCAACGCATGCCAAGTCATCACCTTACCAGAAAGAAACAGAACCCCTCGGCCAG 169
Db 21 SerAlaThrHisAlaLysSerSerProTyrGlnLysThrGluAsnProCysAlaGln 40
QY 170 AGGTGCTCCAGAGTTGTCAACAGAACCGGATGACTTGAAGCAAAAGGCATCGAGTCT 229
Db 41 ArgCysLeuGlnSerCysGlnGlnGluProAspAspLeuLysGlnLysAlaCysGluSer 60
QY 230 CGCTGCACCAAGCTCGAGTATGATCCTCGTTGTGTCTATGATCCTCGAGACACACTGGC 289
Db 61 ArgCysThrLysLeuGluTyrAspProArgCysValTyrAspProArgGlyHisThrGly 80
QY 290 ACCACCAACCAAGTTCCCTCCAGGGGAGCGGACACGTCGGCCGCAACCCCGAGACTAC 349
Db 81 ThrThrAsnGlnArgSerProProGlyGluArgThrArgGlyArgGlnProGlyAspTyr 100
QY 350 GATGATGACCGCGCTCAACCCCAAGAGAGGAGGCGCGATGGGGACCCAGCTGGACCG 409
Db 101 AspAspAspArgGlnProArgArgGluGluGlyArgTyrGlyProAlaGlyPro 120
QY 410 AGGAGCGTGAAAGAGAAAGACTGGAGACCAACCAAGAAAGATTGGAGCGCAAGT 469
Db 121 ArgGluArgGluArgGluAspTyrArgGlnProArgGluAspTyrArgArgProSer 140
QY 470 CATCAGCAGCCAGGAAATTAAGCCCGAAGAGAGAGGAGAAAGCAAGAGTGGGAGACA 529
Db 141 HisGlnGlnProArgLysIleArgProGluGlyArgGluGluGluGluGluThr 160
QY 530 CCAGGTAGCCATCTGAGGGAAGAAAATCTCGGAACAACCTTTCTACTTTCCCGTCAAGG 589
Db 161 ProGlySerHisValArgGluThrSerArgAsnAsnProPheTyrPheProSerArg 180
QY 590 CGGTTTAGCACCGCTTACCGGAAACCAAAACGGTAGGATCCGGGTCTCGAGAGGTTTGAC 649
Db 181 ArgPheSerThrArgTyrGlyAsnGlnAsnGlyArgIleArgValLeuGlnArgPheAsp 200
QY 650 CAAAGGTCAAGCAGTTTCCAGATCTCCAGATCCAGATCCAGATCTTGTGAGATCCAGGCAAA 709
Db 201 GlnArgSerArgGlnPheGlnAsnLeuGlnAsnHisArgIleValGlnIleGluAlaLys 220
QY 710 CTTAACACTCTTGTCTTCTCCCAAGCACGCTGATGCTGATAAATCTTGTATCCAGCAA 769
Db 221 ProAsnThrLeuValLeuProLysHisAlaAspAlaAspAsnIleLeuValIleGlnGln 240
QY 770 GGGCAAGCACCGTGACCGTAGCAATGGCAATGCAATTAACAGAAAGAGCTTTAATCTTGACGAG 829
Db 241 GlyGlnAlaThrValThrValAlaAsnGlyAsnAsnArgLysSerPheAsnLeuLeuAspGlu 260
QY 830 GGCCATGCACTCAGATCCATCCCGGTTTCATTTCTCTACATCTTGAACCGCCATGACAC 889
Db 261 GlyHisAlaLeuArgIleProSerGlyPheIleSerTyrIleLeuAsnArgHisAspAsn 280
QY 890 CAGAACTCAGAGTAGTAAATCTCCATGCCCGTGTAAACACACCCCGGCGAGTTTGAGGAT 949
Db 281 GlnAsnLeuArgAlaAlaLysIleSerMetProValAsnThrProGlyGlnPheGluAsp 300

```
QY 950 TTCTTCCCGCGAGCAGCCGAGACCAATCATCTTCTGAGGGCTTCAGCAGGAATACG 1009
Db 301 PhePheProAlaSerSerArgAspGlnSerSerTyrLeuGlnGlyPheSerArgAsnThr 320
QY 1010 TTGGAGCGCCCTTCAATCGGAATTCATAGATACGAGGGTCTGTTAGAGAGAAAT 1069
Db 321 LeuGluAlaAlaPheAsnAlaGluPheAsnGluIleArgArgValLeuLeuGluAsn 340
QY 1070 GCAGGAGGTGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1129
Db 341 AlaGlyGlyGlnGlnGluArgGlyGlnArgArgTyrSerThrArgSerSerGluAsn 360
QY 1130 AATGAAGGAGTGTAGTCAAGGTGTCAAGAGACACGTTGAAGAACTTACTAAGCACGCT 1189
Db 361 AsnGluGlyValIleValIleValIleValIleValIleValIleValIleValIle 380
QY 1190 AATCCGTCTCAAGAAAGAGCTCCGAGAGAGAGAGAGATATACCAACCAATCAACTTG 1249
Db 381 LysSerValSerLysLysGlySerGluGluGluGlyAspIleThrAsnProIleAsnLeu 400
QY 1250 AGAAGAGCGAGCCGATCTTCTCAACACTTTGGGAGGTATTTCGAGGTGAAGCCAGAC 1309
Db 401 ArgGluGlyGluProAspLeuSerAsnAsnPheGlyLysLeuPheGluValIleVal 420
QY 1310 AAGAAGAACCCAGCTTCAGGACCTCGACATGATGCTCACCTGTGTAGAGATCAAGAA 1369
Db 421 LysLysAsnProGlnLeuGlnAspLeuAspMetMetLeuThrCysValGluIleLysGlu 440
QY 1370 GGAGCTTTGATGTCCACACTTCAAGGCGATGTTATCTGCTCGTCAACAA 1429
Db 441 GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValValVal 460
QY 1430 GGAACCTGAAACCTTGAACTCGTGGCTGTAAAGAAAGAGCAACACAGAGGGGACGGCG 1489
Db 461 GlyThrGlyAsnLeuGluLeuValAlaValArgLysGluGlnGlnArgGlyArgArg 480
QY 1490 GAAGAAGAGGAGGACGAGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1549
Db 481 GluGluGluGluAspGluAspGluGluGluGluGlySerAsnArgGluValArgArgTyr 500
QY 1550 ACAGCGAGTTGAGGAGGAGGAGTGTCTTATCATGCGACAGCTATCCAGTAGCCATC 1609
Db 501 ThrAlaArgLeuLysGluGlyAspValPheIleMetProAlaAlaHisProValAlaIle 520
QY 1610 AACGCTTCTCCGAACTCCATCTGCTTGGCTTCGGTATCAACGCTGAAACCAACACAGA 1669
Db 521 AsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsnAsnHisArg 540
QY 1670 ATCTTCTTGGAGTGATAGGACAATGTGATAGACAGATAGAGAGCAAGCGAGGAT 1729
Db 541 IlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGlnAlaLysAsp 560
QY 1730 TTAGCATTCCTGCTGGTGGGTGAACAAGTGTGAGAAGCTCATCAAAAACACAGAGGAATCT 1789
Db 561 LeuAlaPheProGlySerGlyGlnValGlnValGluLysLeuIleLysAsnGlnLysGluSer 580
QY 1790 CACTTTGTGAGTGTCTGCTCAATCTCAATCTCAATCTCGTCTCGTCTCTGAGAAGAG 1849
Db 581 HisPheValSerAlaArgProGlnSerGlnSerGlnSerProSerSerProGluLysGlu 600
QY 1850 TCTCTGAGAGAGGATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1909
Db 601 SerProGluLysGluAspGlnGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 620
QY 1910 ATTTTGAAGGCTTTTAAAC 1927
Db 621 IleLeuLysAlaPheAsn 626
RESULT 6
ID ABUS2570
XX ABUS2570 standard; protein; 626 AA.
AC ABUS2570;
```

```
XX 10-MAR-2003 (first entry)
DT Peanut Ara h1 mutant V72A.
DE
XX Allergy; Ara h1; Ara h2; Ara h3; IgE binding site; peanut; mutant;
KW mutetin; anaphylactic food allergen; antiallergenic; vaccine;
KW wound healing.
XX Homo sapiens.
XX WO200274250-A2.
XX 26-SEP-2002.
XX 18-MAR-2002; 2002WO-US009108.
XX 16-MAR-2001; 2001US-0276822P.
XX 18-MAR-2002; 2002US-00276822.
XX (PANA-) PANACEA PHARM.
XX Caplan M, Sosin H, Sampson H, Bannan GA, Burks WA, Cockrell G;
PI Compadre CM, Connaughton C, Helm RM, King NE, Kopper RA, Maleki SJ;
PI Rabin PA, Shin DS, Stanley JS;
XX WPI; 2003-018765/01.
XX New modified anaphylactic food allergen, useful for preventing or
PT treating allergic reactions associated with e.g. anaphylactic allergens.
XX Example 5; Page; 30pp; English.
XX The invention relates to a modified anaphylactic food allergen has an
CC amino acid sequence that is substantially identical to that of natural
CC anaphylactic food allergen, except for a cysteine residue that has been
CC modified so that it cannot participate in the disulphide bond. The
CC modification may also comprise mutation of the IgE binding sites to
CC reduce allergenicity. Also included are: (1) a method of making a
CC modified anaphylactic food allergen; (2) a nucleotide molecule encoding
CC or for causing a site specific mutation in the modified anaphylactic food
CC allergen; (3) a transgenic plant or animal expressing the modified
CC anaphylactic food allergen; (4) a method of treating an individual by
CC reducing the clinical response to a natural anaphylactic food allergen;
CC and an isolated fragment of peanut allergen Ara h 1. The modified
CC anaphylactic food allergen is useful for preventing or treating allergic
CC reactions associated with any natural allergen such as food, insect,
CC rubber or preferably anaphylactic allergens. It is also useful for
CC treating wounds in mammals such as bovine, canine, feline, caprine,
CC ovine, porcine, murine or equine species. The present sequence is a food
CC allergen, mutated to alter its IgE binding characteristics. Note: The
CC present sequence is not shown in the specification but was created by the
CC indexer using information provided in the specification
XX
XX SQ Sequence 626 AA;
Alignment Scores: 3 38e-304 Length: 626
Pred. No.: 3282.00 Matches: 625
Percent Similarity: 99.84% Conservative: 0
Best Local Similarity: 99.84% Mismatches: 1
Query Match: 92.14% Indels: 0
DB: 6 Gaps: 0
US-10-728-051-1 (1-2032) x ABUS2570 (1-626)
QY 50 ATGAGAGGAGGGTTCCTCCACTGATGCTGCTAGGATCCTTCTGCTTCAGTT 109
Db 1 MetArgGlyArgValSerProLeuMetLeuLeuGlyIleLeuValLeuAlaSerVal 20
QY 110 TCTGCAACGATCGCAAGTTCATCATCTTACCAAGAAAAACAGAAACCCCTGCGCCAG 169
Db 21 SerAlaThrHisAlaLysSerSerProTyrGlnLysLysThrGluAsnProCysAlaGln 40
```

```
QY 170 AGGTGCTCCAGAGTTGTCAACAGGAACCGGATGACTTGAGCAAAAGGATCGCGAGTCT 229
DB 41 ArgCysLeuGlnSerCysGlnGlnGluProAspAspLeuLysGlnLysAlaCysGluSer 60
QY 230 CGCTGCACCAAGCTCCAGATGATGATCTCGTTGTCTATCATCTCTCGAGGACACACTGGC 289
DB 61 ArgCysThrLysLeuGluLysAspProArgCysAlaLysAspProArgGlyHisThrGly 80
QY 290 ACCACCAACCAAGTTCCCTCCAGGGGAGCGACACAGTGGCGCCCAACCCCGAGACTAC 349
DB 81 ThrThrAsnGlnArgSerProProGlyGluArgThrArgGlyArgGlnProGlyAspTyr 100
QY 350 GATGATGACCGCGCTCAACCCGAAAGAGAGGAAGCGCGATGGGGACAGCTGAGACCG 409
DB 101 AspAspAspArgGlnProArgArgGluGluGlyGlyArgTrpGlyProAlaGlyPro 120
QY 410 AGGGCGGTGAAGACAGAAGACTCGAGACACCAAGAGAGATGGAGGGCGACCAACT 469
DB 121 ArgGluArgGluArgGluAspTrpArgGlnProArgGluAspTrpArgProSer 140
QY 470 CATCAGCAGCCAGGAAATAAGCCCGAAGAGAGAGAAACAAGAGTGGGGAACA 529
DB 141 HisGlnGlnProArgLysIleArgProGluGlyArgGluGlyGluGlnTrpGlyThr 160
QY 530 CAGGTAGCCATGTGAGGAAGAAAATCTCGGAACAACCTTTCTACTTCCCGTCAAGG 589
DB 161 ProGlySerHisValArgGluGluThrSerArgAsnAsnProPheTyrPheProSerArg 180
QY 590 CGGTTTAGCACCGCTACCGGACCAACAAACGGTAGGATCGGCTCTCGAGAGTTTGAC 649
DB 181 ArgPheSerThrArgTyrGlyAsnGlnAsnGlyArgIleArgValLeuGlnArgPheAsp 200
QY 650 CAAAGCTCAAGGAGTTTCAGAAATCCAGAAATCACCGTATTGTGCAGATCCAGGCGCAA 709
DB 201 GlnArgSerArgGlnPheGlnAsnLeuGlnAsnHisArgIleValGlnIleGluAlaLys 220
QY 710 CTAAACACTCTGTTCTCCAGACAGCTGATGCTGATAAATCTTGTATTCAGCAA 769
DB 221 ProAsnThrLeuValLeuProLysHisAlaAspAlaAspAsnIleLeuValIleGlnGln 240
QY 770 GGGCAAGCCACCTGACCTAGCAATGCCAATGCAATACAGAAAGCTTTAATCTTGACGAG 829
DB 241 GlyGlnAlaThrValThrValAlaAsnGlyAsnAsnArgLysSerPheAsnLeuAspGlu 260
QY 830 GGCATGCACTCAGAAATCCCATCCGGTTTCATTTCTACATCTTGAAACCGCCATGACAAC 889
DB 261 GlyHisAlaLeuArgIleProSerGlyPheIleSerTyrIleLeuAsnArgHisAspAsn 280
QY 890 CAGAACTCAGAGTAGCTAAATCTCCATGCCCGTTAACACACCCCGCAGTTTGAGGAT 949
DB 281 GlnAsnLeuArgValAlaLysIleSerMetProValAsnThrProGlyGlnPheGluAsp 300
QY 950 TTCTTCCCGCGAGCGGAGCCGACCAATCATCTACTTCAGGGCTTCAGCAGGATACG 1009
DB 301 PheProAlaSerSerArgAspGlnSerSerTyrLeuGlnGlyPheSerArgAsnThr 320
QY 1010 TTGGAGCGCGCTTCAATCGGAATTCATAGATACGAGGAGTGTGTGTAGAGAGAAT 1069
DB 321 LeuGluAlaAlaPheAsnAlaGluPheAsnGluIleArgArgValLeuLeuGluGluAsn 340
QY 1070 GCAGGAGGTGACAGAGGAGAGAGCGGACAGCGGATGGAGTACTCGGAGTAGTGAGAAC 1129
DB 341 AlaGlyGlyGluGlnGluArgGlyGlnArgArgTrpSerThrArgSerSerGluAsn 360
QY 1130 AATGAGGAGTAGTAGTCAAGTGTCAAGGACACAGTTGAGAACTTACTAAGCAGCT 1189
DB 361 AsnGluGlyValIleValLysValSerLysGluHisValGluGluLeuThrLysHisAla 380
QY 1190 AAATCGTCTCAAGAAAGCTCCGAGAGAGAGGAGATATCACCAACCAATCAACTTG 1249
DB 381 LysSerValSerLysGlySerGluGluGluGlyAspIleThrAsnProIleAsnLeu 400
```

```
QY 1250 AGAGAAGGCGAGCCCGATCTTTCAACAACTTTGGAAAGTTATTTGAGGTGAAGCCAGAC 1309
DB 401 ArgGluGlyGluProAspLeuSerAsnAsnPheGlyLysLeuPheGluValLysProAsp 420
QY 1310 AAGAAGAACCCCGAGCTTCAGACCTCGGACATGATGCTCACCTGTGTAGAGATCAAGAA 1369
DB 421 LysLysAsnProGlnLeuGlnAspLeuAspMetLeuThrCysValGluIleLysGlu 440
QY 1370 GGAGCTTTGATGCTCCACACTTCACTCAAAAGGCCATGGTTATCTCGTCTGCTCAACAA 1429
DB 441 GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValAsnLys 460
QY 1430 GGAATCGGAAACCTTGAACCTCGTGGCTGTAAAGAAAAGACAAACAAGAGGGGACGGCG 1489
DB 461 GlyThrGlyAsnLeuGluValAlaValArgLysGluGlnGlnArgGlyArgArg 480
QY 1490 GAAGAAGAGGAGGACGACGACGAAGAGAGGAGGAAGTAACAGAGAGGTGCGTAGGTAC 1549
DB 481 GluGluGluGluAspGluAspGluGluGluGlySerAsnArgGluValArgArgTyr 500
QY 1550 ACAGCGAGGTTGAAGGAAGCGGATGTGTTTCATCATGCCAGCAGCTCATCCAGTAGCCATC 1609
DB 501 ThrAlaArgLeuLysGluGlyAspValPheIleMetProAlaAlaHisProValAlaIle 520
QY 1610 AAGCTTCTCCGAACCTCCATCTGCTTGGCTTCGGTATCAACGCTGAAACCAACACAGA 1669
DB 521 AsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsnAsnHisArg 540
QY 1670 ATCTTCTTGCAGGTGATAAGGACCAATGTAGTAGACCATGAGAGCAAGCAAGCAAGGAT 1729
DB 541 IlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGlnAlaLysAsp 560
QY 1730 TTAGCATTCCTCGGTCGGGTGAACAAAGTTGAGAAGCTCATCAAAAACCCAGAGGAATCT 1789
DB 561 LeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGlnLysGluSer 580
QY 1790 CACTTTGTGAGTGTGCTCCTCAATCTCAATCTCAATCTCCGTCGCTCTCTGAGAAAGAG 1849
DB 581 HisPheValSerAlaArgProGlnSerGlnSerGlnSerProSerSerProGluLysGlu 600
QY 1850 TCTCTCGAAGAGAGGATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1909
DB 601 SerProGluLysGluAspGlnGluGluAsnGlnGlyLysGlyProLeuLeuSer 620
QY 1910 ATTTTGAAGGCTTTTAAAC 1927
DB 621 IleLeuLysAlaPheAsn 626
RESULT 7
ABU52568
ID ABU52568 standard; protein; 626 AA.
XX
AC ABU52568;
XX
DT 10-MAR-2003 (first entry)
XX
DE Peanut Ara h1 mutant K32A.
XX
KW Allergy; Ara h1; Ara h2; Ara h3; IgE binding site; peanut; mutant;
KW mutein; anaphylactic food allergen; anti-allergenic; vaccine;
XX wound healing.
XX
OS Homo sapiens.
XX
PN WO200274250-A2.
XX
PD 26-SEP-2002.
XX
PF 18-MAR-2002; 2002WO-US009108.
XX
PR 16-MAR-2001; 2001US-0276822P.
PR 18-MAR-2002; 2002US-00276822.
XX
```


Best Local Similarity: 99.84%		Mismatches: 1
Query Match: 92.08%	Indels: 0	
DB: 6	Gaps: 0	
US-10-728-051-1 (1-2032) x ABUS2574 (1-626)		
QY	50	ATCAGAGGAGGGTTTCTCCACTGATGCTGTCTAGGATCCTTGTCTCGGCTTCAGTT 109
DB	1	MetArgIYArgValSerProLeuMetLeuLeuGlyLeuValLeuAlaSerVal 20
QY	110	TCGCAACGATCCCAAGTCATCACCTTACCAGAAGAAACAGAACCCCTCGGCCAG 169
DB	21	SerAlaThrHisAlaYssSerProTyrGlnYssThrGluAsnProCysAlaGln 40
QY	170	AGTGCTCCAGATTGTCAACAGGAACCGGATGACTTGAAGCAAAAGGATCGAGTCT 229
DB	41	ArgCysLeuGlnSerCysGlnGlnProAspAspLeuYssGlnYssAlaCysGluSer 60
QY	230	CGCTGCACCAAGTCAGATATGATCTCTGTGTGTCTATGATCTCTCGAGGACACTGGC 289
DB	61	ArgCysThrIysLeuGluTyrAspProArgCysValTyrAspProArgGlyHisThrGly 80
QY	290	ACCACCAACCAAGTTCCTCCAGGGAGCGACACGTCGGCCCAACCCCGAGACTAC 349
DB	81	ThrThrAsnGlnArgSerProProGlyGluArgThrArgGlyArgGlnProGlyAspTyr 100
QY	350	GATGATGACCGCGTCAACCCCAAGAGAGAGAGCGGATGGGACCCAGCTGGACCG 409
DB	101	AspAspArgArgGlnProArgGluArgGluGlyArgTrpGlyProAlaGlyPro 120
QY	410	AGGAGCGTGAAGAGAGAAGACTGGAGACAACCAAGAGAAGATTGGAGGCGCAAGT 469
DB	121	ArgGluArgGluArgGluGluAspTrpArgGlnProArgGluAspTrpArgGlnProSer 140
QY	470	CATCAGCAGCCAGGAATAGCCCGGAGGAGAGAGGAGAAACAAGAGTGGGAAACA 529
DB	141	HisGlnGlnProArgIysIleArgProGlyArgGlyGluGlnGluTrpGlyThr 160
QY	530	CCAGGTAGCATGTAGGGAAGAAACATCTCGGAACAACCTTTCTACTTCCCGTCAAGG 589
DB	161	ProGlySerHisValArgGluGluThrSerArgAsnAsnProPheTyrPheProSerArg 180
QY	590	CGTTTAGCACCCGCTACGGGAACCAAAACGGTAGGATCCGGTCTCTGAGAGTTTGAC 649
DB	181	ArgPheSerThrArgTyrGlyAsnGlnAsnGlyArgIleArgValLeuGlnArgPheAsp 200
QY	650	CRAAGCTCAAGGAGTTCAGATCTCCAGAAATCCAGATTCAGATTCGAGATCGAGCCAA 709
DB	201	GlnArgSerArgGlnPheGlnAsnLeuGlnAsnHisArgIleValGlnIleGluAlaLys 220
QY	710	CCTAACACTTCTTCTCCAGCAGCTGATGCTGATAACATCTTGTATCCAGCAA 769
DB	221	ProAsnThrLeuValLeuProLysHisAlaAspAlaAspAsnIleLeuValIleGlnGln 240
QY	770	GGCAAGCCACCGTAGCCGTAAGTAATGGCAATAACAGAAAGCTTTAATCTTGACGAG 829
DB	241	GlyGlnAlaThrValThrValAlaAsnGlyAsnAsnArgIysSerPheAsnLeuAspGlu 260
QY	830	GGCATGCACTCAGATCCCATCCGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 889
DB	261	GlyHisAlaLeuArgIleProSerGlyPheIleSerTyrIleLeuAsnArgHisAspAsn 280
QY	890	CAGAACCTCAGATGAGTAAATCTCCATCCCGTAAACACACCCCGCCAGTTTGGAGAT 949
DB	281	GlnAsnLeuArgValAlaLysIleSerMetProValAsnThrProGlyGlnPheGluAsp 300
QY	950	TTCTTCCCGCGAGCCGAGACCAATCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1009
DB	301	PhePheProAlaSerSerArgAspGlnSerSerTyrLeuGlnGlyPheSerArgAsnThr 320
QY	1010	TTGAGGCCCGCTTCAATCGGAATTCATAGATACGAGGAGGTGCTGTTAGAGAGAT 1069
DB	321	LeuGluAlaAlaPheAsnAlaGluPheAsnGluIleArgArgValLeuLeuGluAsn 340
QY	1070	GCAGAGGTGAGCAAGAGAGAGAGAGGCGAGCGGATGAGTACTCGAGTACTGAGAAC 1129
DB	341	AlaGlyGlyGluGlnGluGluArgGlyGlnArgArgTrpSerThrArgSerSerGluAsn 360
QY	1130	AATGAGGAGTGTAGTCAAGTGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1189
DB	361	AsnGluGlyValIleValIysValSerLysGluHisValGluGluLeuThrLysHisAla 380
QY	1190	AAATCGCTCTCAAGAAAGGCTCCGAAGAGAGGAGAGATATCACCAACCAATCAACTTG 1249
DB	381	LysSerValSerLysLysGlySerGluGluGluGlyAspIleThrAsnProIleAsnLeu 400
QY	1250	AGAGAAGCGAGCCCGATCTTTCTAACTTTGGGAAGTATTATTGAGTGAAGCCAGAC 1309
DB	401	ArgGluGlyGluProAspLeuSerAsnAsnPheGlyLysLeuPheGluValLysProAsp 420
QY	1310	AGAGAAGCCCGAGCTTCAGGACCTGGACATGATGCTCACCTGTCTAGAGATCAAGAA 1369
DB	421	LysLysAsnProGlnLeuGlnAspLeuAspMetLeuThrCysValGluIleLysGlu 440
QY	1370	CGAGCTTTGATGCTCCACACTTCAACTCAAGGCCATGTTATCTCGTCTGCTCAACAA 1429
DB	441	GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValValAsnLys 460
QY	1430	GGAATCGGAACTTGAACCTCGTGTCTGTAAGAAAAGACAAACAAGAGGAGCGCGG 1489
DB	461	GlyThrGlyAsnLeuGluLeuValAlaValArgLysGluGlnGlnArgGlyArgArg 480
QY	1490	GAAGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1549
DB	481	GluGluGluGluAspGluAspGluGluGluGlySerAsnArgGluValArgAlaTyr 500
QY	1550	ACAGCGAGTTCGAAGAGCGGATGTTTCATCATCCAGCAGCTCATCCAGTAGGCATC 1609
DB	501	ThrAlaArgLeuLysGlyAspValPheIleMetProAlaHisProValAlaIle 520
QY	1610	AACGTTCTCCGAACTCCATCTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCT 1669
DB	521	AsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsnHisArg 540
QY	1670	ATCTTCTTCCAGGTGATAGGACATGTGTAGACCATAGACATAGACATAGACATAGACAT 1729
DB	541	IlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGlnAlaLysAsp 560
QY	1730	TTAGCATTCCTGGGTGGGTGAAACAAGTTGAGAAGCTCATCAAAACAGAGAGGAATCT 1789
DB	561	LeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGlnLysGluSer 580
QY	1790	CACTTTGTGAGTCTCTCTCAATCTCAATCTCAATCTCAATCTCAATCTCAATCTCAATCT 1849
DB	581	HisPheValSerAlaArgProGlnSerGlnSerGlnSerProSerSerProGluLysGlu 600
QY	1850	TCTCTGAGAAAGAGATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1909
DB	601	SerProGluLysGluAspGlnGluGluGluAsnGlnGlyGlyLysGlyProLeuLeuSer 620
QY	1910	ATTTTGAAGGCTTTTAAAC 1927
DB	621	IleLeuLysAlaPheAsn 626
RESULT 10		
ABUS2573		
ID	ABUS2573	standard; protein; 626 AA.
XX	ABUS2573;	
DT	10-MAR-2003	(first entry)
XX	Peanut Ara h1 mutant R109A.	
XX	Allergy; Ara h1; Ara h2; Ara h3; IgE binding site; peanut; mutant;	
KW	mucin; anaphylactic food allergen; anti-allergenic; vaccine;	

KW wound healing.
 OS Homo sapiens.
 PN WO200274250-A2.
 XX 26-SEP-2002.
 PD 18-MAR-2002; 2002WO-US009108.
 PF 16-MAR-2001; 2001US-0276822P.
 PR 18-MAR-2002; 2002US-00276822.
 XX (PANA-) PANACEA PHARM.
 XX Caplan M, Sosin H, Sampson H, Bannan GA, Burks WA, Cockrell G;
 PI Compadre CM, Connaughton C, Helm RM, King NE, Kopper RA, Maleki SJ;
 PI Rabjohn PA, Shin DS, Stanley JS;
 XX WPI; 2003-018765/01.
 XX New modified anaphylactic food allergen, useful for preventing or
 PT treating allergic reactions associated with e.g. anaphylactic allergens.
 PT
 XX Example 5; Page; 300pp; English.
 XX The invention relates to a modified anaphylactic food allergen has an
 CC amino acid sequence that is substantially identical to that of natural
 CC anaphylactic food allergen, except for a cysteine residue that has been
 CC modified so that it cannot participate in the disulphide bond. The
 CC modification may also comprise mutation of the Igs binding sites to
 CC reduce allergenicity. Also included are: (1) a method of making a
 CC modified anaphylactic food allergen; (2) a nucleotide molecule encoding
 CC or for causing a site specific mutation in the modified anaphylactic food
 CC allergen; (3) a transgenic plant or animal expressing the modified
 CC anaphylactic food allergen; (4) a method of treating an individual by
 CC reducing the clinical response to a natural anaphylactic food allergen;
 CC and an isolated fragment of peanut allergen Ara h 1. The modified
 CC anaphylactic food allergen is useful for preventing or treating allergic
 CC reactions associated with any natural allergen such as food, insect,
 CC rubber or preferably anaphylactic allergens. It is also useful for
 CC treating wounds in mammals such as bovine, canine, feline, caprine,
 CC ovine, porcine, murine or equine species. The present sequence is a food
 CC allergen, mutated to alter its IGE binding characteristics. Note: The
 CC present sequence is not shown in the specification but was created by the
 CC indexer using information provided in the specification
 XX
 SQ Sequence 626 AA;

 Alignment Scores:
 Pred. No.: 5-25e-304 Length: 626
 Score: 3280.00 Matches: 625
 Percent Similarity: 99.84% Conservative: 0
 Best Local Similarity: 99.84% Mismatches: 1
 Query Match: 92.08% Indels: 0
 DB: 6 Gaps: 0

 US-10-728-051-1 (1-2032) x ABUS2573 (1-626)
 QY 50 ATGAGAGGGGGTTCCTCCACTGATGCTTGCTAGGATCCTTGCTGGCTTCAGTT 109
 DB 1 MetArgGlyArgValSerProLeuMetLeuLeuLeuGlyLeuValLeuAlaSerVal 20
 QY 110 TCTGCAACGCATGCCAAGTCATCACCTTACCAAGAAAAACAGAAACCCCTGCGCCAG 169
 DB 21 SerAlaThrHisAlaLysSerSerProTyrGlnLysLysThrGluAsnProCysAlaGln 40
 QY 170 AGTGCTCCAGATTGTCAACAGGAACCGGATGACTTGAAGCAAAAGCATGCGAGTCT 229
 DB 41 ArgCysLeuGlnSerCysGlnGlnGluProAspAspLeuLysGlnLysAlaCysGluSer 60
 QY 230 CGCTGCACCAAGCTCCAGTATGATCTCGTTGTGCTATGATCCTCGAGGACACACTGGC 289

Db 61 ArgCysThrLysLeuGluTyrAspProArgCysValTyrAspProArgGlyHisThrGly 80
 QY 290 ACCACCAACCAACCTTCCCTCCAGGGAGCGACACAGTGGCGCCCAACCCGAGACTAC 349
 DB 81 ThrThrAsnGlnArgSerProProGlyGluArgThrArgGlyArgGlnProGlyAspTyr 100
 QY 350 GATGATGACCCGCGTCAACCCCGAAGAGAGAGAGAGCGCGATGGGGACCACTGGACCG 409
 DB 101 AspAspArgArgGlnProArgAlaGluGluGlyGlyArgTrpGlyProAlaGlyPro 120
 QY 410 AGGGAGCGTGAAGAGAGAGAACTGGAGACAAACCAAGAGAGATTGGAGGCGCAACT 469
 DB 121 ArgGluArgGluArgGluAspTrpArgGlnProArgGluAspTrpArgProSer 140
 QY 470 CATCAGACCCACGAGGAAATAAGGCCGGAAGAGAGAGAGAGAAACAAAGTGGGGAACA 529
 DB 141 HisGlnGlnProArgLysIleArgProGluGlyArgGluGlyGluGlnGlnTrpGlyThr 160
 QY 530 CCAGGTAGCCATGTGAGGGAAGAAAATCTCGAAACAAACCCCTTCTACTCCCGTCAAGG 589
 DB 161 ProGlySerHisValArgGluThrSerArgAsnAsnProPheTyrPheProSerArg 180
 QY 590 CGGTTTAGCCCGCTACGGGAACCAAAACGCTAGGATCCGGGTCTCGCAGAGTTGAC 649
 DB 181 ArgPheSerThrArgTyrGlyAsnGlnAsnGlyArgIleArgValLeuGlnArgPheAsp 200
 QY 650 CAAAGTCAAGCAGTTCCTCAGAAATCTCAGAAATCACCGTATTGTGCAGATCCAGGCCAAA 709
 DB 201 GlnArgSerArgGlnPheGlnAsnLeuGlnAsnHisArgIleValGlnIleGluAlaLys 220
 QY 710 CTTAACTCTTGTCTTCCCAAGCAGCTGATGTGATAACATCCTCTTGTATCCAGCAA 769
 DB 221 ProAsnThrLeuValLeuProLysHisAlaAspAlaAspAsnIleLeuValIleGlnGln 240
 QY 770 GGGCAGCCACCGCTGACGCTAGCAATGCGCATACAGAAAGAGCTTAACTCTTCAGCAG 829
 DB 241 GlyGlnAlaThrValThrValAlaAsnGlyAsnAsnArgLysSerPheAsnLeuAspGlu 260
 QY 830 GGCCATGCACTCAGAAATCCATCCCGTTCATTTCTTCTACATCTTGAACCGCATCACAAC 889
 DB 261 GlyHisAlaLeuArgIleProSerGlyPheIleSerTyrIleLeuAsnArgHisAspAsn 280
 QY 890 CAGAACCTCAGAGTAGCTAAATCTCCATGCGCCGTTAACAACCCGCGCAGTTTGGAGAT 949
 DB 281 GlnAsnLeuArgValAlaLysIleSerMetProValAsnThrProGlyGlnPheGluAsp 300
 QY 950 TTTCTTCCGCGCAGCAGCGACCAATCATCTACTTGCAGGCTTTCAGCAGGATACG 1009
 DB 301 PhePheProAlaSerSerArgAspGlnSerSerTyrLeuGlnGlyPheSerArgAsnThr 320
 QY 1010 TTGGAGGCGCCCTTCAATGCGGAATTCATGAGATACGAGGAGTGTGTTAGAAAGAGAT 1069
 DB 321 LeuGluAlaAlaPheAsnAlaGluPheAsnGluIleArgArgValLeuLeuGluAsn 340
 QY 1070 GCAGGAGTGACCAAGAGAGAGAGCGCAGCGCATGGAGTACTCGGAGTAGTGAGAAC 1129
 DB 341 AlaGlyGlyGluGlnGluArgGlyGlnArgGlyGlnArgTrpSerThrArgSerSerGluAsn 360
 QY 1130 AATGAGGAGTGCATAGTCAAAGTGTCAAAGGAGCAGCTTGAAGAACTTACTAAGCACGCT 1189
 DB 361 AsnGluGlyValIleValLysValSerLysGluHisValGluGluLeuThrLysHisAla 380
 QY 1190 AAATCCGCTCTCAAGAAAGGCTCCGGAAGAGAGGAGATATACCAACCCCAATCACTTG 1249
 DB 381 LysSerValSerLysLysGlySerGluGluGlyAspIleThrAsnProIleAsnLeu 400
 QY 1250 AGAGAAGCGAGCCCGATCTTTCTAACTTTGGGAAGTTATTTGAGGTGAAGCCAGAC 1309
 DB 401 ArgGluGlyGluProAspLeuSerAsnAsnPheGlyLysLeuPheGluValLysProAsp 420
 QY 1310 AAGAGAAGACCCCGCTTCAGGACCTGGACATGATGTCTCACCTGTGTGAGATCAAGAA 1369
 DB 421 LysLysAsnProGlnLeuGlnAspLeuAspMetMetLeuThrCysValGluLysGlu 440

```
QY 1370 GAGCTTTGATGCTCCACACTTCACTCACTAAAGGCCATGTTATCGTCGTCGCAACAAA 1429
|||
Db 441 GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValValAsnLys 460
|||
QY 1430 GGAAGTGGAAACCTTGAACTCGTGGCTGTGAAGAAAGAGCAACAACAGAGGGGACGGCGG 1489
|||
Db 461 GlyThrGlyAsnLeuGluLeuValAlaValArgLysGluGlnGlnArgGlyArgArg 480
|||
QY 1490 GAAGAAGAGGAGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1549
|||
Db 481 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 500
|||
QY 1550 ACAGCGAGGTGAAGGAGCGATGTTTCATATGTCACGACGACGACGACGACGACGACGAC 1609
|||
Db 501 ThrAlaArgLeuLysGluGlyArgValPheIleMetProAlaAlaHisProValAlaIle 520
|||
QY 1610 AACGCTTCCCTCCGAACTCCATCTGCTGGCTTCGGTATCAACGCTGAACCAACCAACCA 1669
|||
Db 521 AsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsnHisArg 540
|||
QY 1670 ATCTTCTTCAGGTGATAAGGACAATGTGATAGACGACGACGACGACGACGACGACGACG 1729
|||
Db 541 IlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGlnAlaLysAsp 560
|||
QY 1730 TTAGCATTCCTGGTGGGTGAACAAGTTGAGAAGCTCATCAAAACCAAGAGGAATCT 1789
|||
Db 561 LeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGlnLysGluSer 580
|||
QY 1790 CACTTGTGAGTCTGCTCCTCAATCTCAATCTCAATCTCGCTCGCTCCTCGTGAAGAG 1849
|||
Db 581 HisPheValSerAlaArgProGlnSerGlnSerGlnSerProSerProGluLysGlu 600
|||
QY 1850 TCTCCTCAGAAAGAGGATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1909
|||
Db 601 SerProGluLysGluAspGlnGluGluAsnGlnGlyGlyLysGlyProLeuLeuSer 620
|||
QY 1910 ATTTTGAAGCGTTTAAAC 1927
|||
Db 621 IleLeuLysAlaPheAsn 626
|||
RESULT 11
ADM12053
ID ADM12053 standard; protein; 626 AA.
XX AC
XX ADM12053;
XX DT
XX 20-MAY-2004 (first entry)
XX DE
XX Arachis hypogaea 1 (Ara hi) protein.
XX KW
XX antigen presentation enhancing hybrid polypeptide; mammalian Ii-Key;
XX MHC class II; antibacterial; virucide; fungicide; antirheumatic;
XX antiarthritic; neuroprotective; dermatological; immunosuppressive;
XX antiinflammatory; antidiabetic; antithyroid; immune;
XX rheumatoid arthritis; multiple sclerosis; lupus erythematosus;
XX diabetes mellitus; myasthenia gravis; autoimmune thyroiditis;
XX scleroderma; dermatomyositis; pemphigus.
XX OS
XX Arachis hypogaea.
XX PN
XX US2003235594-A1.
XX XX
XX 25-DEC-2003.
XX XX
XX 17-SEP-2002; 2002US-00245871.
XX PF
XX
XX 14-SEP-1999; 99US-00396813.
XX PR
XX 17-JUL-2002; 2002US-00197000.
XX XX
XX (ANTI-) ANTIGEN EXPRESS INC.
XX XX
XX Humphreys R, Xu M;
```

XX WPI; 2004-070554/07.

XX Novel Ii-Key/antigen presentation enhancing hybrid polypeptide, useful
PT for treating infections, rheumatoid arthritis, multiple sclerosis, lupus
PT erythematosus and diabetes mellitus.

XX Example 1; Page 16-17; 87pp; English.

XX The invention relates to a novel antigen presentation enhancing hybrid
CC polypeptide. The novel polypeptide has an N-terminal element consisting
CC of 4-16 residues of a mammalian Ii-Key peptide and its non-N-terminal
CC deletion modifications, a chemical structure covalently linking the N-
CC terminal element to an MHC class II-presented epitope of a C-terminal
CC element. The C-terminal element comprises an antigenic epitope, which
CC binds to an antigenic peptide binding site of an MHC class II molecule.
CC The antigen presentation enhancing hybrid polypeptide has the following
CC activities: antibacterial, virucide, fungicide, antirheumatic,
CC antiarthritic, neuroprotective, dermatological, immunosuppressive,
CC antiinflammatory, antidiabetic, and antithyroid. The antigen presentation
CC enhancing hybrid polypeptide is useful for modulating the immune response
CC in an individual and for treating infections (such as bacteria, virus,
CC parasite and fungus), rheumatoid arthritis, multiple sclerosis, lupus
CC erythematosus, diabetes mellitus, myasthenia gravis, autoimmune
CC thyroiditis, scleroderma, dermatomyositis and pemphigus. This sequence
CC represents a mammalian Ii key related protein of the invention.

XX SQ Sequence 626 AA;

Alignment Scores:
Pred. No.: 6,54e-304 Length: 626
Score: 3279.00 Matches: 625
Percent Similarity: 99.84% Conservative: 0
Best Local Similarity: 99.84% Mismatches: 1
Query Match: 92.06% Indels: 0
DB: 8 Gaps: 0

US-10-728-051-1 (1-2032) x ADM12053 (1-626)

```
QY 50 ATGAGAGGGAGGGTTTCTCCACTGATGCTGTTCTAGGGATCCTTGCTCGGCTTCAGTT 109
|||
Db 1 MetArgLysValSerProLeuMetLeuLeuGlyIleLeuValLeuAlaSerVal 20
|||
QY 110 TCTGCAACGCATGCCAAGTCATCACCTTACCAGAGAAAACAGAGAACCCCTCGGCCAG 169
|||
Db 21 SerAlaThrHisAlaLysSerSerProTyrGlnLysLysThrGluAsnProCysAlaGln 40
|||
QY 170 AGTGCCTCCAGATGTTGTCACAGAACCGGATGACTTGAGCAAAAGGATCGGAGTCT 229
|||
Db 41 ArgCysLeuGlnSerCysGlnGlnGluProAspAspLeuLysGlyLysAlaCysGluSer 60
|||
QY 230 CGCTGCACCAAGCTCCGAGTATGATCCTCGTTGTCTATGATCCTCGAGGACACACTGGC 289
|||
Db 61 ArgCysThrLysLeuGluTyrAspProArgCysValTyrAspProArgGlyHisThrGly 80
|||
QY 290 ACCACCAACCAACGTTCCCTCCAGGGGACGACACGTCGCGCCCAACCCGAGACTAC 349
|||
Db 81 ThrThrAsnGlnArgSerProGlyGluArgThrArgGlyArgGlnProGlyAspTyr 100
|||
QY 350 GATGATGACCGCGCTCAACCCCGAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 409
|||
Db 101 AspAspAspArgArgGlnProArgGluGluGlyArgTyrGlyProAlaGlyPro 120
|||
QY 410 AGGAGCGCTGAAAGAGAAGACTGGAGACAACCAAGAGAAGATTGGAGCGCACCAAGT 469
|||
Db 121 ArgGluArgGluArgGluGluAspTyrAspGlnProArgGluAspTyrAspArgProSer 140
|||
QY 470 CATCAGCAGCCACGAAAAATAAGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 529
|||
Db 141 HisGlnGlnProArgLysIleArgProGluGlyArgGluGlyGluGlnGluTyrGlyThr 160
|||
QY 530 CCAGGTAGCCATGTGAGGGAGAGAAACATCTCGGAACACCCCTTCTACTTCCCGTCAAG 589
|||
```

161 ProGlySerHisValArgGluThrSerArgAsnAsnProPheTyrPheProSerArg 180
162 ProGlySerHisValArgGluThrSerArgAsnAsnProPheTyrPheProSerArg 180
590 CGGTTTACACCGCTTACCGGACCAACAAACGCTAGGATCCCGGTCCTCGACAGGTTTGAC 649
181 ArgPheSerThrArgTyrGlyAsnGlnAsnGlyArgValLeuGlnArgPheAsp 200
650 CAAAGTCAGGCGAGTTTCAGAAATCTCCAGATACACCGTATGTCAGATCCAGGCCAAA 709
201 GlnArgSerArgGlnPheGlnAsnLeuGlnAsnHisArgIleValGlnIleGluAlaLys 220
710 CTTAAACACTCTGTTCTCCCAAGCAGCTGATGCTGATAACATCTGTTATTCAGCAAA 769
221 ProAsnThrLeuValLeuProLysHisAlaAspAlaAspAsnIleLeuValIleGlnGln 240
770 GGCACAGCCACCGTCAGCGGTAGCAATGCAATGCAATGCAATGCAATGCAATGCAATG 829
241 GlyGlnAlaThrValThrValAlaAsnGlyAsnAsnArgLysSerPheAsnLeuAspGlu 260
830 GGCATGCACTCAAGATCCCATCCGCTTCAATTCCTTACATCTTGTAACCGCATGACAA 889
261 GlyHisAlaLeuArgIleProSerGlyPheIleSerTyrIleLeuAsnArgHisAspAsn 280
890 CAGAACCTCAGAGTAGTAAATCTCCATGCCGTTTAAACACACCCCGCCAGTTTGAGGAT 949
281 GlnAsnLeuArgValAlaLysIleSerMetProValAsnThrProGlyGlnPheGluAsp 300
950 TTCTTCCCGCGAGCAGCCGAGACCAATCATCTTACTTGCAGGGCTTCAGCAGGATACG 1009
301 PhePheProAlaSerSerArgAspGlnSerTyrLeuGlnGlyPheSerArgAsnThr 320
1010 TTGAGGCGCGCTTCAATCGGGAATTCATGAGATACGAGGCTGCTGTAGAGAGAAAT 1069
321 LeuGluAlaAlaPheAsnAlaGluPheAsnGluIleArgArgValLeuLeuGluAsn 340
1070 GCAGAGGTGACGAGAGGAGGAGGCGAGGCGATGGAGTACTCGAGTAGTAGAAC 1129
341 AlaGlyGlyGluGlnGluArgGlyGlnArgArgTrpSerThrArgSerSerGluAsn 360
1130 AATGAGAGGTAGTACGTAAGTCTCAAGTCTCAAGGAGCAGCTTGAAGAACTTACTAAGACGCT 1189
361 AsnGluGlyValIleValLysValSerLysGluHisValGluGluLeuThrLysHisAla 380
1190 AAATCCGCTCTCAAAGAAGGCTCCGAAGAAGAGGAGATATACCAACCCCAATCACTTG 1249
381 LysSerValSerLysLysGlySerGluGluGluGlyAspIleThrAsnProIleAsnLeu 400
1250 AGAGAGGCGAGCCGATCTTCTTAACTTTGGAGTATTGAGGTGAAGCCAGAC 1309
401 ArgGluGlyGluProAspLeuSerAsnAsnPheGlyLysLeuPheGluValLysProAsp 420
1310 AAGAGAACCCCGCAGCTTCAGACCTGGACATGCTCACCTGCTGAGAGATCAAGAA 1369
421 LysLysAsnProGlnLeuGlnAspLeuAspMetMetLeuThrCysValGluIleLysGlu 440
1370 GGAGCTTTGATGCTCCACATTCAACTCAAGGCGATGTTATCGTCTGCTCAACAAA 1429
441 GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValValAsnLys 460
1430 GGAACCTGGAACCTTGAATCTGCTGCTGTAAGAAAGAGCAACAGAGGGGACGGCG 1489
461 GlyThrGlyAsnLeuGluLeuValAlaValArgLysGluGlnGlnArgGlyArgArg 480
1490 GAAGAGAGGAGGAGCAAGACCAAGAGAGGAGGAGGAGTAAACAGAGAGTGGTAGTAC 1549
481 GluGluGluGluAspGluGluGluGluGlySerAsnArgGluValArgArgTyr 500
1550 ACAGCAGGTGAGAGGCGATGTTTCATCATGCCAGAGCTCATCCAGTACCCATC 1609
501 ThrAlaArgLeuLysGluGlyAspValPheIleMetProAlaAlaHisProValAlaIle 520
1610 AACGCTTCTCCGAACTCATCTGCTTGGCTTCCGTTATCAACGCTGAAACCAACACAGA 1669
521 AsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsnHisArg 540

1670 ATCTTCTCTGAGTGTATAGGACAAATGTATAGACCATAGACAGCAACGCAAGGAT 1729
541 IlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGlnAlaLysAsp 560
1730 TTACGATTCCTCGGTCGGGTGAACAGTTGAGAGCTCATCAAAACAGAGGAATCT 1789
561 LeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGlnLysGluSer 580
1790 CACTTTGTGAGTCTGCTCTCAATCTCAATCTCAATCTCCGTCGCTCTCTGAGAAAGAG 1849
581 HisPheValSerAlaArgProGlnSerGlnSerGlnSerProSerSerProGluLysGlu 600
1850 TCTCTCAGAAAGAGGATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1909
601 SerProGluLysGluAspGlnGluGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 620
1910 ATTTTGAAGCTTTTAAAC 1927
621 IleLeuLysAlaPheAsn 626
RESULT 12
ABU52569 standard; protein; 626 AA.
ID ABU52569 standard; protein; 626 AA.
AC ABU52569;
XX 10-MAR-2003 (first entry)
DT Peanut Ara h1 mutant D52A.
DE Allergy; Ara h1; Ara h2; Ara h3; IgE binding site; peanut; mutant;
KW mutin; anaphylactic food allergen; antiallergenic; vaccine;
KW wound healing.
XX Homo sapiens.
OS WO200274250-A2.
XX 26-SEP-2002.
XX 18-MAR-2002; 2002WO-US009108.
XX 16-MAR-2001; 2001US-0276822P.
XX 18-MAR-2002; 2002US-00276822.
XX (PANA-) PANACEA PHARM.
XX Caplan M, Sosin H, Sampson H, Bannan GA, Burks WA, Cockrell G;
XX Compadre CM, Connaughton C, Helm RM, King NE, Kopper RA, Maleki SJ;
XX Rabjohn PA, Shin DS, Stanley JS;
XX WPI; 2003-018765/01.
XX New modified anaphylactic food allergen, useful for preventing or
XX treating allergic reactions associated with e.g. anaphylactic allergens.
XX Example 5; Page; 300pp; English.
XX The invention relates to a modified anaphylactic food allergen has an
XX amino acid sequence that is substantially identical to that of natural
XX anaphylactic food allergen, except for a cysteine residue that has been
XX modified so that it cannot participate in the disulphide bond. The
XX modification may also comprise mutation of the IgE binding sites to
XX reduce allergenicity. Also included are: (1) a method of making a
XX modified anaphylactic food allergen; (2) a nucleotide molecule encoding
XX or for causing a site specific mutation in the modified anaphylactic food
XX allergen; (3) a transgenic plant or animal expressing the modified
XX anaphylactic food allergen; (4) a method of treating an individual by
XX reducing the clinical response to a natural anaphylactic food allergen;
XX and an isolated fragment of peanut allergen Ara h 1. The modified
XX anaphylactic food allergen is useful for preventing or treating allergic
XX reactions associated with any natural allergen such as food, insect,

CC rubber or preferably anaphylactic allergens. It is also useful for
CC treating wounds in mammals such as bovine, canine, feline, caprine,
CC ovine, porcine, murine or equine species. The present sequence is a food
CC allergen, mutated to alter its IgE binding characteristics. Note: The
CC present sequence is not shown in the specification but was created by the
CC indexer using information provided in the specification

XX
SQ Sequence 626 AA;

Alignment Scores:

Pred. No.: 8.15e-304 Length: 626
Score: 3278.00 Matches: 625
Percent Similarity: 99.84% Conservative: 0
Best Local Similarity: 99.84% Mismatches: 1
Query Match: 92.03% Indels: 0
DB: 6 Gaps: 0

US-10-728-051-1 (1-2032) x ABUS2569 (1-626)

```
QY 50 ATGAGAGGAGGGTTTCTCCACGTGATGCTGTGCTAGGATCTCTGTGCTGGCTTCAGTT 109
DB 1 MetArgGlyArgValSerProLeuMetLeuLeuGlyIleLeuValLeuAlaSerVal 20
QY 110 TCTGCAACGATGCCAAGTCATCACCTTACCAGAAGAAACACAGAACCCCTCGGCCAG 169
DB 21 SerAlaThrHisAlaLysSerProTyrGlnLysLysThrGluAsnProCysAlaGln 40
QY 170 AGGTGCTCCAGAGTTGTCAACAGGAACCGGATGACTTGAAGCAAAAGGATCGAGTCT 229
DB 41 ArgCysLeuGlnSerCysGlnGlnProAspAlaLeuLysGlnLysAlaCysGluSer 60
QY 230 CGCTGCAACCAAGTCGAGTATGATCTCTGTTGTCTATGATCTCGAGACACACTGGC 289
DB 61 ArgCysThrLysLeuGluTyrAspProArgCysValTyrAspProArgGlyHisThrGly 80
QY 290 ACCACCAACCAAGTTCCTCCAGGGAGCGGACACGTCGCGCCCAACCCGAGACTAC 349
DB 81 ThrThrAsnGlnArgSerProProGlyGluArgThrArgGlyArgGlnProGlyAspTyr 100
QY 350 GATGATACCCCGCTCAACCCCGAAGAGAGAGAGCGGCGATGGGACACCAAGTGGACCG 409
DB 101 AspAspArgArgGlnProArgArgGluGluGlyArgTrpGlyProAlaGlyPro 120
QY 410 AGGAGCGTGAAGAGAAAGACTGGAGACACCAAGAGAAAGATTGGAGCGGACCAAGT 469
DB 121 ArgGluArgGluArgGluAspTrpArgGlnProArgGluAspTrpArgProSer 140
QY 470 CATCAGCAGCCAGGAAATAGCCCGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 529
DB 141 HisGlnProArgLysIleArgProGluGlyArgGluGluGlnGluTrpGlyThr 160
QY 530 CCAGGTAGCCATGTGAGGGAAGAAACATCTCGGAACCAACCTTTCTACTTCCCGTCAAGG 589
DB 161 ProGlySerHisValArgGluGluThrSerArgAsnAsnProPheTyrPheProSerArg 180
QY 590 CGGTTTACGACCCGCTACGGGAACCAAAACGGTAGGATCCGGGTCTGCGAGAGTTTGAC 649
DB 181 ArgPheSerThrArgTyrGlyAsnGlnAsnGlyArgIleArgValLeuGlnArgPheAsp 200
QY 650 CAAAGGTCAAGGAGTTTTCAGATCTCCAGATCACCGTATCTGTCAGATCGAGGCCAAA 709
DB 201 GlnArgSerArgGlnPheGlnAsnLeuGlnAsnHisArgIleValGlnIleGluAlaLys 220
QY 710 CCTAACACTCTTGTCTTCCCAAGCAGCGTGATGCTGATTAACATCCCTTGTATTCCAGCAA 769
DB 221 ProAsnThrLeuValLeuProLysHisAlaAspAlaAspAsnIleLeuValIleGlnGln 240
QY 770 GGGCAAGCCCGTAGCCGTAGCAAAATGGCAATTAACAGAAAGAGCTTTAATCTTTGACGAG 829
DB 241 GlyGlnAlaThrValThrValAlaAsnGlyAsnAsnArgLysSerPheAsnLeuAspGlu 260
QY 830 GGCCATGCACACTCAGAAATCCCATCCGGTTTCAATTCCTACATCTTGAACCCGCTACCA 889
```

```
DB 261 GlyHisAlaLeuArgIleProSerGlyPheIleSerTyrIleLeuAsnArgHisAspAsn 280
QY 890 CAGAACTCAGAGTAGCTAAATCTCCATGCGCGTTAACACACCCGCGCATTTGAGGAT 949
DB 281 GlnAsnLeuArgValAlaLysIleSerMetProValAsnThrProGlyGlnPheGluAsp 300
QY 950 TTCTTCCCGGAGCAGCCGAGACCAATCATCTACTTCCAGGGCTTCAGCAGGATACG 1009
DB 301 PhePheProAlaSerSerArgAspGlnSerSerTyrLeuGlnGlyPheSerArgAsnThr 320
QY 1010 TTGAGGCGCGCTTCAATCGGAATTCATAGATACGAGGAGGTGCTGTTAGAGAGAAT 1069
DB 321 LeuGluAlaAlaPheAsnAlaGluPheAsnGluIleArgArgValLeuLeuGluAsn 340
QY 1070 GCAGAGGTGAGCAAGAGGAGAGAGCGGAGAGCGATGGAGTACTCGGAGTAGTGAAC 1129
DB 341 AlaGlyGlyGluGlnGluArgGlyGlnArgArgTrpSerThrArgSerSerGluAsn 360
QY 1130 AATGAGGAGTGTAGTCAAGTGTCAAGGAGCAGGTTGAAGAACTTACTTAGACGCT 1189
DB 361 AsnGluGlyValIleValLysValSerLysGluHisValGluGluLeuThrLysHisAla 380
QY 1190 AATCCGTCTCAAAGAAAGGCTCCGAAGAGAGGAGATATCACCAACCCCAATCAACTTG 1249
DB 381 LysSerValSerLysLysGlySerGluGluGlyAspIleThrAsnProIleAsnLeu 400
QY 1250 AGAAGGCGAGCCGATCTTTCTAACTTTGGGAAGTTATTTGAGGTGAAGCCAGAC 1309
DB 401 ArgGluGlyGluProAspLeuSerAsnAsnPheGlyLysLeuPheGluValLysProAsp 420
QY 1310 AAGAAACCCCGAGCTTCAGGACCTGGACATGATCTACCTGTGTAGAGATCAAGAA 1369
DB 421 LysLysAsnProGlnLeuGlnAspLeuAspMetMetLeuThrCysValGluIleLysGlu 440
QY 1370 GGAGCTTTGATGCTCCACACTTCAACTCAAAGCCATGTTATCGTCGTCGCAACAAA 1429
DB 441 GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValAsnLys 460
QY 1430 GSACTGGAAACCTTCAACTCGTGGCTGAAGAAAAGAGCAACACAGAGGGAGCGCGG 1489
DB 461 GlyThrGlyAsnLeuGluValAlaValArgLysGluGlnGlnArgGlyArg 480
QY 1490 GAAAGAGGAGGACCAAGACGAGAGAGAGGAGGAGTAACAGAGAGGTGCGCTAGTAC 1549
DB 481 GluGluGluGluAspGluGluGluGluGlySerAsnArgGluValArgTyr 500
QY 1550 ACAGCGAGGTGAAGGAAGCGGATGTTTCATCATGCGCAGCTCATCCAGTAGCCATC 1609
DB 501 ThrAlaArgLeuLysGluGlyAspValPheIleMetProAlaHisProValAlaIle 520
QY 1610 AACGTTCTCTCCGAATCCATCTGCTTGGCTTCGGTATCAACCTGAAACCAACCCAGAG 1669
DB 521 AsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsnAsnHisArg 540
QY 1670 ATCTTCTTCAGGTCATTAAGCAATGTGATAGACAGATAGAGAGAGCGAGGAT 1729
DB 541 IlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGlnAlaLysAsp 560
QY 1730 TTAGCATTTCCCTCGGTGAACTGAAGTTGAGAAGCTCATCAAAACCAAGAGGAATCT 1789
DB 561 LeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGlnLysGluSer 580
QY 1790 CACTTTGTGAGTCTCGTCTCAATCTCAATCTCAATCTCCGCTGCTCTCTGAGAAAGAG 1849
DB 581 HisPheValSerAlaArgProGlnSerGlnSerProSerProSerProGluLysGlu 600
QY 1850 TCTCTGAGAAAGAGATCAAGAGGAGGAGAAACCAAGGAGGAGGAGGTCCTCCTCTTCA 1909
DB 601 SerProGluLysGluAspGlnGluGluAsnGlnGlyGlyLysGlyProLeuLeuSer 620
QY 1910 ATTTTGAAGGCTTTTAAAC 1927
DB 621 IleLeuLysAlaPheAsn 626
```

RESULT 13

ABU52572
ID ABU52572 standard; protein; 626 AA.XX AC
XX AC
XX AC
XX AC

DT 10-MAR-2003 (first entry)

XX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DEXX DE
XX DE
XX DE
XX DE

1 MetArgGlyArgValSerProLeuMetLeuLeuGlyIleLeuValLeuAlaSerVal 20
110 TCTCAACGCGATGCCAAGTATCATCTACCTTACCAGAGAGAAACAGAGAACCCCTGGCCCGAG 169
21 SerAlaThrHisAlaLysSerProTyrGlnLysThrGluAsnProCysAlaGln 40
170 AGGTGCTCTCCAGAGTTCCTCAACAGAGAACCGGATGACTTGAAGCAAAAGGATGCGAGTCT 229
41 ArgCysLeuGlnSerCysGlnGluProAspAspLeuLysGlnLysAlaCysGluSer 60
230 CGGTGACCAAGCTCGAGTATGATCTCTGCTGCTATGATCTCTGAGGACACACTGGC 289
61 ArgCysThrLysLeuGluTyrAspProArgCysValTyrAspProArgGlyHisThrGly 80
290 ACCACCAACCAACGTTCCCTCCAGGAGGAGGACGTCGCGCCCAACCCGAGAGACTAC 349
81 ThrThrAsnGlnArgSerProGlyGluArgThrArgGlyArgGlnProGlyAspTyr 100
350 GATGATGACCGCCGCTCAACCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 409
101 AspAspAlaArgArgGlnProArgArgGluGlyArgGlyArgGlyProAlaGlyPro 120
410 AGGAGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 469
121 ArgGluArgGluArgGluAspTrpArgGlnProArgGluAspTrpArgGluAspTrpArg 140
470 CATCAGCAGCCAGCAAAATTAAGGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 529
141 HisGlnGlnProArgGlyArgGlnGluGluGluGluGluGluGluGluGluGluGluGlu 160
530 CCAGGTAGCCATGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 589
161 ProGlySerHisValArgGluGluThrSerArgAsnAsnProPheTyrPheProSerArg 180
590 CGGTTAGCAGCCGCTACGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 649
181 ArgPheSerThrArgTyrGlyAsnGlnAsnGlyArgGluGluGluGluGluGluGluGlu 200
650 CAAAGGTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 709
201 GlnArgSerArgGlnPheGlnAsnLeuGlnAsnHisArgIleValGlnIleGluAlaLys 220
710 CTTACACTCTTGTCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 769
221 ProAsnThrLeuValLeuProLysHisAlaAspAlaAspAsnIleLeuValIleGlnGln 240
770 GGGCAGCCAGCAGTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 829
241 GlyGlnAlaThrValThrValAlaAsnGlyAsnAsnArgLysSerPheAsnLeuAspGlu 260
830 GGGCAGTCACTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 889
261 GlyHisAlaLeuArgIleProSerGlyPheIleSerTyrIleLeuAsnArgHisAspAsn 280
890 CAGAACCTCAGAGTACGAGTAAATCTCCATGCGCGCTTACACACACCGCGAGTTTCAGGAT 949
281 GlnAsnLeuArgValAlaLysIleSerMetProValAsnThrProGlyGlnPheGluAsp 300
950 TTTCTCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1009
301 PhePheProAlaSerSerArgAspGlnSerSerTyrLeuGlnGlyPheSerArgAsnThr 320
1010 TTGGAGCGCCCTTCAATCGGAAATTCATAGATACGAGAGAGAGAGAGAGAGAGAGAGAG 1069
321 LeuGluAlaAlaPheAsnAlaGluPheAsnGluIleArgArgValLeuLeuGluGluAsn 340
1070 GCAGGAGGTGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1129
341 AlaGlyGlyGluGlnGluGluArgGlyGlnArgArgIleSerThrArgSerSerGluAsn 360
1130 AATGAAGGAGTGTATGATCAAGTGTCAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1189

US-10-728-051-1 (1-2032) x ABU52572 (1-626)

50 ATGAGAGGAGGTTTCTCCACTGATGCTGTGTAGGAGTCTTGTCTGCTTCTAGTT 109

```
Db 361 AsnGluGlyValIleValIleValIleValSerLysGluHisValGluGluLeuThrLysHisAla 390
Qy 1190 AAATCCGCTCTCAAGAAGAGCTCCGAGAGAGAGATATACCAACCAATCAACTTG 1249
Db 381 LysSerValSerLysLysGlySerGluGluGluGlyAspIleThrAsnProIleAsnLeu 400
Qy 1250 AGAAGAGGCGAGCCGATCTTTCTAACAACCTTTGGGAAGTTATTTGAGGTGAAGCCAGAC 1309
Db 401 ArgGluGlyGluProAspLeuSerAsnAsnPheGlyLysLeuPheGluValLysProAsp 420
Qy 1310 AAGAAGAACCCCGAGCTTCAGGACCTCGACATGATGCTACCTGTGTAGAGATCAAGAA 1369
Db 421 LysLysAsnProGlnLeuGlnAspLeuAspMetLeuThrCysValGluIleLysGlu 440
Qy 1370 GGAGCTTTGATGCTCCACACTTCAACTCAAGGCCATGTTATCGTCGTCGTCACAA 1429
Db 441 GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValAsnLys 460
Qy 1430 GGAAGTGAACCTTGAACCTCGCTGCTGTAAGAAAGAGCAACAACAGAGGCGACGGCG 1489
Db 461 GlyThrGlyAsnLeuGluLeuAlaValArgLysGluGlnGlnArgGlyArgArg 480
Qy 1490 GAAGAAGAGGAGGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1549
Db 481 GluGluGluGluAspGluAspGluGluGluGluGluGluGluGluGluValArgArgTyr 500
Qy 1550 ACAGCGAGGTGAAGAGAGCGATGTGTTTCATCATGCCAGCTCATCCAGTAGCCATC 1609
Db 501 ThrAlaArgLeuLysGluGlyAspValPheIleMetProAlaAlaHisProValAlaIle 520
Qy 1610 AACGCTTCCTCCGAATCCATCTGCTTGGCTTCGTTATCAACCTGAAACCAACACAGA 1669
Db 521 AsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsnAsnHisArg 540
Qy 1670 ATCTTCTCTCAGGTATAGGACAATGTGATAGCCAGATAGAGCAAGCAAGCAAGGAT 1729
Db 541 IlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGlnAlaLysAsp 560
Qy 1730 TTAGCATTCCTCGGTCGGTCAACAGTTGAGAAGCTCATCAAAACCAAGAGGAATCT 1789
Db 561 LeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGlnLysGluSer 580
Qy 1790 CACTTTGTGAGTCTGCTCTCAATCTCAATCTCAATCTCGTCTCTCTGAGAAAGAG 1849
Db 581 HisPheValSerAlaArgProGlnSerGlnSerGlnSerProSerSerProGluLysGlu 600
Qy 1850 TCTCTCTGAGAAAGAGATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1909
Db 601 SerProGluLysGluAspGlnGlnGluGluAsnGlnGlyLysGlyProLeuLeuSer 620
Qy 1910 ATTTTGAAGGCTTTTAAAC 1927
Db 621 IleLeuLysAlaPheAsn 626

RESULT 14
AAW22150
ID AAW22150 standard; protein; 626 AA.
AC AAW22150;
XX AAW22150;
DT 17-OCT-2003 (revised)
DT 29-DEC-1997 (first entry)
XX XX
DE Peanut allergen Ara hi.
XX Peanut; seed storage protein; allergen; allergy; hypersensitivity;
XX vaccine; anaphylactic shock; immunotherapy; therapy; monoclonal antibody;
XX ELISA; analysis; Ara hi.
XX Arachis hypogaea; strain Flotrunner.
XX Key Location/Qualifiers
FT Peptide 1..22
```

```
FT Protein /label= Sig_peptide
FT 23..626
FT /label= Mat_protein
FT Modified-site 521..523
FT FT /note= "N-glycosylation site"
XX
PN WO9724139-Al.
XX 10-JUL-1997.
XX 23-SEP-1996; 96WO-US015222.
XX 29-DEC-1995; 95US-0009455P.
PR 04-MAR-1996; 96US-00610424.
XX (UYAR-) UNIV ARKANSAS.
XX
PI Burks AW, Helm RM, Cockrell G, Stanley JS, Bannon GA;
XX WPI; 1997-363453/33.
DR N-PSDB; AAT76613.
XX
PT Peanut allergens Ara hi and Ara hII - used for vaccination and in two-
XX site monoclonal antibody based ELISA.
PS Claim 31; Page 172; 354pp; English.
XX
CC This polypeptide comprises major peanut allergen Ara hi (AAW22149). Its
CC sequence was deduced from cDNA clone p41b (AAT76613), isolated from
CC peanut seed cDNA using a primer (see AAT76616) based on an isolated Ara
CC hi peptide (see AAW24206). The sequence shows significant homology with
CC the vicilin family of seed storage proteins of other legumes. The
CC allergen is recognised by serum IGE from a large proportion of
CC individuals with peanut hypersensitivity. Ara hi and Ara hII (see
CC AAW24164) can be used to raise monoclonal antibodies which are used in a
CC specific two-site Mab ELISA for the detection of Ara hi or Ara hII
CC (claimed). IGE-binding Ara hi antigen epitopes (see AAW24165-87) may be
CC used in vaccines to protect against allergic reactions to peanut
CC allergens, e.g. anaphylactic shock. (Updated on 17-OCT-2003 to
CC standardise OS field)
XX
SQ Sequence 626 AA;
```

Alignment Scores:

```
Pred. No.: 1..97e-303 Length: 626
Score: 3274.00 Matches: 624
Percent Similarity: 99.68% Conservative: 0
Best Local Similarity: 99.68% Mismatches: 2
Query Match: 91.91% Indels: 0
DB: 2 Gaps: 0
```

US-10-728-051-1 (1-2032) x AAW22150 (1-626)

```
Qy 50 ATGAGAGGAGGGGTTTCTCCACTGATGCTTCTAGGATCTCTGCTGCTTCAAGTT 109
Db 1 MetArgGlyArgValSerProLeuMetLeuLeuLeuGlyIleLeuValLeuAlaSerVal 20
Qy 110 TCTGCAACGCATGCCAAGTCATCACCTTACCAAGAAAGAAACAGAGAACCCCTCGCCAG 169
Db 21 SerAlaThrHisAlaLysSerSerProTyrGlnLysLysThrGluAsnProCysAlaGln 40
Qy 170 AGGTGCTCCAGAGTTGTCAACAGGAACCGGATGACTTGAAGCAAAAGGATCGAGTCT 229
Db 41 ArgCysLeuGlnSerCysGlnGlnGluProAspAspLeuLysGlnLysAlaCysGluSer 60
Qy 230 CGCTGCACCAAGCTCCAGTATGATCTCGTTGTGTCTATGATCTCGAGGACACACTGGC 289
Db 61 ArgCysThrLysLeuGluTyrAspProArgCysValTyrAspProArgGlyHisThrGly 80
Qy 290 ACCACCAACCAAGCTTCCCTCCAGGGGAGCGACACAGTGGCGGCCCAACCCGGAGACTAC 349
Db 81 ThrThrAsnGlnArgSerProGlyGluArgThrArgGlyArgGlnProGlyAspTyr 100
```


QY 350 GATGATGACCGCGCTCAACCCCGAGAGAGAGGCGGATGGGACCACTGGACG 409
Db 101 AspAspAspArgGlnProArgArgGluGluGlyGlyArgTrpGlyProAlaGlyPro 120
QY 410 AGGAGCGGTGAAGAGAGAAAGACTCGACACAACCAAGAGAAAGATTGGAGGCGCAAGT 469
Db 121 ArgGluArgGluArgGluAspTrpArgGlnProArgGluAspTrpArgArgProSer 140
QY 470 CATCAGCAGCCACGGAATAAGCCCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 529
Db 141 HisGlnGlnProArgLysIleArgProGluGlyArgGluGlyGluGlnGluTrpGlyThr 160
QY 530 CAGGTAGCCATGTGAGGAGAGAAATCTCCGAAACAACCCCTTCTACCTCCGCTCAAG 589
Db 161 ProGlySerHisValArgGluGluThrSerArgAsnProPheTrpSerArg 180
QY 590 CGGTTTAGCACCGCTACGGGAACCAAAACGGTAGGATCCGGTCTCGAGAGGTTTGAC 649
Db 181 ArgPheSerThrArgTrpGlyAsnGlnAsnGlyArgValLeuGlnArgPheAsp 200
QY 650 CAAAGGTCAAGGAGGTTTCAAGATCTCCAGAAATCACCGTATTGTGCAGATCCAGGCCAA 709
Db 201 GlnArgSerArgGlnPheGlnAsnLeuGlnAsnHisArgIleValGlnIleGluAlaLys 220
QY 710 CCTAACACTCTGTCTCTCCAAAGCAGCTGATGCTGATAACATCTCTTGTATCCAGCAA 769
Db 221 ProAsnThrLeuValLeuPheLysHisAlaAspAlaAspAsnIleLeuValIleGln 240
QY 770 GGGCAAGCCCGTGACCGGTAGCAAAATGGCAATTAACAGAAAGAGCTTTTAACTCTTGACGAG 829
Db 241 GlyGlnAlaThrValThrValAlaAsnGlyAsnAsnArgLysSerPheAsnLeuAspGlu 260
QY 830 GGCCATCGACTCAGAACCCATCCGTTTCAATCTCTTACATCTTGAAACCGCCATGACAAC 889
Db 261 GlyHisAlaLeuArgIleProSerGlyPheIleSerTrpIleLeuAsnArgHisAspAsn 280
QY 890 CAGAACCTCAGAGTAGTAAATCTCCATGCCGTTTAAACACACCCGCCCGCTTTGAGGAT 949
Db 281 GlnAsnLeuArgValAlaLysIleSerMetProValThrGlnProGlyGlnPheGluAsp 300
QY 950 TTCTTCCCGGAGCAGCGAGAGCAATCATCTCTACTTGTGAGGCTTACAGCGATACG 1009
Db 301 PhePheProAlaSerSerArgAspGlnSerSerTrpLeuGlnGlyPheSerArgAsnThr 320
QY 1010 TTGGAGCGCGCTTCAATCGGAATTCATGAGATACGGAGGTCGTGTAGAGAGAAAT 1069
Db 321 LeuGluAlaAlaPheAsnAlaGluPheAsnGluIleArgValLeuLeuGluGluAsn 340
QY 1070 GCAGGAGGTGAGCAGAGAGAGAGCGGATGGAGTACTCGGAGTAGTGAGAAC 1129
Db 341 AlaGlyGlyGluGlnGluArgGlyGlnArgArgTrpSerThrArgSerSerGluAsn 360
QY 1130 AATGAGAGGTAGTAGTCAAGTGTCAAGAGAGCAGCTTGAAGAACTTACTAAGACGCT 1189
Db 361 AsnGluGlyValIleValLysValSerLysGluHisValGluLeuLeuThrLysHisAla 380
QY 1190 AAATCCGCTCAAGAAAGGCTCCGAGAGAGAGAGAGATATCACCAACCCCAATCAACTTG 1249
Db 381 LysSerValSerLysLysGlySerGluGluGluGlyAspIleThrAsnProIleAsnLeu 400
QY 1250 AGAGAGCGGAGCGCGATCTTTCTTAACTTTGGGAAGTTATTGAGGTGAGCCAGAC 1309
Db 401 ArgGluGlyGluProAspLeuSerAsnAsnPheGlyLysLeuPheGluValLysProAsp 420
QY 1310 RAGAGAACCCCGCTTCCAGACCTGACATGATGCTCACTGTGTAGAGATCAAGAGAA 1369
Db 421 LysLysAsnProGlnLeuGlnAspLeuAspMetMetLeuThrCysValGluIleLysGlu 440
QY 1370 GGAGCTTTGATGCTCCCACTTCAACTCAAAGGCCATGGTTATCGTCTCGTCAACAAA 1429
Db 441 GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValValAsnLys 460
QY 1430 GGAACCTGGNAACCTTGAACTCTGGCTGTAGAAAAGAGCAACACAGAGGGGACGGCG 1489

Db 461 GlyThrGlyAsnLeuGluLeuValAlaValArgLysGluGlnGlnArgGlyArg 480
QY 1490 GAAGAAGAGGAGGACGAGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1549
Db 481 GluGluGluGluAspGluAspGluGluGluGluGlySerAsnArgGluValArgArgTyr 500
QY 1550 ACAGCGAGTTGAAGAGAGCGGATGTCTCATATGCAGCAGCTCATCCAGTAGCCATC 1609
Db 501 ThrAlaArgLeuLysGluGlyAspValPheIleMetProAlaAlaHisProValAlaIle 520
QY 1610 AACGCTTCTCCGAACCTCATCTCTTGGCTTCGGTATCAACGCTGAAAAACAACACAGA 1669
Db 521 AsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsnAsnHisArg 540
QY 1670 ATCTCTCTTCAGCTGATTAAGCAATGTGTATACACAGATAGAGAGAGAGAGAGAGAG 1729
Db 541 IlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGlnAlaLysAsp 560
QY 1730 TTAGCATTCCTCGGTCGGGTGAACAAAGTTGAGAAGCTCATCAAAAACCAAGAGAGATCT 1789
Db 561 LeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGlnLysGluSer 580
QY 1790 CACTTTGTAGTGTCTCTCTCAATCTCAATCTCAATCTCAATCTCAATCTCAATCTCA 1849
Db 581 HisPheValSerAlaArgProGlnSerGlnSerGlnSerProSerSerProGluLysGlu 600
QY 1850 TCTCTCTGAGAAGAGGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1909
Db 601 SerProGluLysGluAspGlnGluGluGlnGluGlnGlyGlyLysGlyProLeuLeuSer 620
QY 1910 ATTTTGAAGGCTTTTAAAC 1927
Db 621 IleLeuLysAlaPheAsn 626
RESULT 15
ADG27464
ID ADG27464 standard; protein; 625 AA.
XX AC ADG27464;
XX DT 26-FEB-2004 (first entry)
XX DE Peanut Ara h1 cDNA clone P41b protein.
XX KW Peanut; plant; allergen; Ara h1; Ara h2; Ara h3; glycinin A2B1a; Jug nl;
XX KW antiallergic; vulnery; anaphylactic food allergen; IgE; allergy; wound.
XX OS Arachis hypogaea.
XX PN US2003202980-A1.
XX PD 30-OCT-2003.
XX PF 18-MAR-2002; 2002US-00100303.
XX PR 29-DEC-1995; 95US-0009455P.
PR 23-SEP-1996; 96US-00717933.
PR 31-JAN-1998; 98US-0073283P.
PR 13-FEB-1998; 98US-0074590P.
PR 13-FEB-1998; 98US-0074624P.
PR 13-FEB-1998; 98US-0074633P.
PR 23-JUN-1998; 98US-00105872.
PR 27-AUG-1998; 98US-00141220.
PR 13-NOV-1998; 98US-00191593.
PR 29-JAN-1999; 99US-00240557.
PR 29-JAN-1999; 99US-00241101.
PR 11-FEB-1999; 99US-00248673.
PR 11-FEB-1999; 99US-00248674.
PR 02-MAR-1999; 99US-0122450P.
PR 02-MAR-1999; 99US-0122452P.
PR 02-MAR-1999; 99US-0122560P.
PR 02-MAR-1999; 99US-0122565P.

Db 361 AsnGluGlyValIleValIysValSerLysGluHieValGluLeuThrLysHisAla 380
QY 1190 AATCCGTCCTCAAGAAAGCTCCGAGAGAGGAGATATCACCAACCAACTCACTTG 1249
Db 381 LysSerValSerLysLysGlySerGluGluGluGlyAspIleThrAsnProIleAsnLeu 400
QY 1250 AGAGAGGCCAGCCGATCTTTCTAACTTTGGGAAGTTATTTGAGGTGAAGCCAGAC 1309
Db 401 ArgGluGlyGluProAspLeuSerAsnAsnAsnPheGlyLysLeuPheGluValLysProAsp 420
QY 1310 AAGAAGAACCCAGCTTCAGGACCTGGACATGCTCACCTGTGTAGAGATCAAGAA 1369
Db 421 LysLysAsnProGlnLeuGlnAspLeuAspMetLeuThrCysValGluIleLysGlu 440
QY 1370 GGAGCTTTGATGTCCTCCACACTTCAACTCAAGCCCATGGTTATCGTCGTCACAAA 1429
Db 441 GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValAsnLys 460
QY 1430 GGAACCTGAAACCTTGAACTCGTGGCTGTAAAGAAAGAGCAACAACAGAGGGGACGCGG 1489
Db 461 GlyThrGlyAsnLeuGluLeuValAlaValArgLysGluGlnGlnArgGlyArg 480
QY 1490 GAAGAAGAGGAGGACGAGAGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1549
Db 481 GluGluGluAspGluAspGluGluGluGluGluGluGluGluGluGluGluGluGlu 500
QY 1550 ACAGCGAGTTGAGAGAGCGATGCTTCATCATGCCAGCAGCTCATCCAGTAGCCATC 1609
Db 501 ThrAlaArgLeuLysGluGlyAspValPheIleMetProAlaAlaHisProValAlaIle 520
QY 1610 AACGCTTCCTCCGAACCTCCATCTCTTGGCTTCGGTATCAACGCTGAAACCAACACACAGA 1669
Db 521 AsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsnAsnHisArg 540
QY 1670 ATCTTCCTTGAGGTGATAGGACAATGTGATAGACCAGATAGAGAGAGAGAGAGAGAGAT 1729
Db 541 IlePheLeuAlaGly---LysAspAsnValIleAspGlnIleGluLysGlnAlaLysAsp 559
QY 1730 TTGCAATTCCTGGGTCCGGTGAACAAGTTGAGAAGCTCATCAAAAACAGAGAGGATCT 1789
Db 560 LeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGlnLysGluSer 579
QY 1790 CACTTTGTGAGTGTCTGCTCAATCTCAATCTCAATCTCGTCTCTCTGAGAAAGAG 1849
Db 580 HisPheValSerAlaArgProGlnSerGlnSerGlnSerProSerSerProGluLysGlu 599
QY 1850 TCTCCTGAGAGAGGATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1909
Db 600 SerProGluLysGluAspGlnGluGluGlnGluGlnGlyLysGlyProLeuLeuSer 619
QY 1910 ATTTTGAAGGCTTTTAAC 1927
Db 620 IleLeuLysAlaPheAsn 625

Search completed: April 8, 2005, 05:01:13
Job time : 336.773 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 8, 2005, 05:25:59 ; Search time 233.254 Seconds
(without alignments)

5784.401 Million cell updates/sec

Title: US-10-728-051-1

Perfect score: 3562

Sequence: 1 aataacatataattcattcgcgtttgtgtttctcc 2032

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1418010 seqs, 331997259 residues

Total number of hits satisfying chosen parameters: 2836020

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2_1/USPTO.spool.p/US10728051/runat_07042005.125048.14128/app.query.fasta_1.4757
-DB=Published Applications AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=DIOSUM62
-TRANS=human40.cdd -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10728051 @CGN 1.1 663 @runat_07042005.125048.14128
-NCPUS=6 -ICPU=3 -NO MMAP -LARGESUBSTR -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

Description

No. Score Match Length DB ID

1	3286	92.3	626	10	US-09-847-208-28	Sequence 28, Appl
2	3286	92.3	626	14	US-10-228-806-2	Sequence 2, Appl
3	3286	92.3	626	15	US-10-100-303A-7	Sequence 7, Appl
4	3286	92.3	626	15	US-10-245-871-10	Sequence 10, Appl
5	3286	92.3	626	15	US-10-253-286-10	Sequence 10, Appl
6	3286	92.3	626	17	US-10-899-551-2	Sequence 2, Appl
7	3192	89.6	635	17	US-10-899-551-53	Sequence 53, Appl
8	3052	85.7	634	9	US-09-731-221-78	Sequence 78, Appl
9	3041	85.4	614	9	US-09-331-631A-21	Sequence 21, Appl
10	3041	85.4	614	10	US-09-847-208-27	Sequence 27, Appl
11	3041	85.4	614	14	US-10-147-095-21	Sequence 21, Appl
12	3041	85.4	614	15	US-10-100-303A-8	Sequence 8, Appl
13	1394	39.1	268	15	US-10-100-303A-55	Sequence 55, Appl
14	1282	36.0	605	15	US-10-424-599-153195	Sequence 153195,
15	1281	36.0	605	9	US-09-331-631A-25	Sequence 25, Appl
16	1281	36.0	605	14	US-10-147-095-25	Sequence 25, Appl
17	1275.5	35.8	623	15	US-10-424-599-153206	Sequence 153206,
18	1268	35.6	605	15	US-10-100-303A-110	Sequence 110, App
19	1201.5	33.7	584	15	US-10-424-599-260105	Sequence 260105,
20	1157	32.5	417	15	US-10-245-227B-14	Sequence 14, Appl
21	1156	32.5	425	15	US-10-245-227B-12	Sequence 12, Appl
22	1156	32.5	439	15	US-10-245-227B-1	Sequence 1, Appl
23	1117.5	31.4	390	15	US-10-245-227B-2	Sequence 2, Appl
24	874	24.5	324	15	US-10-425-114-44408	Sequence 44408, A
25	867	24.3	344	15	US-10-425-114-51703	Sequence 51703, A
26	865.5	24.3	666	9	US-09-331-631A-3	Sequence 3, Appl
27	865.5	24.3	666	14	US-10-147-095-3	Sequence 3, Appl
28	863.5	24.2	625	9	US-09-331-631A-5	Sequence 5, Appl
29	863.5	24.2	625	14	US-10-147-095-5	Sequence 5, Appl
30	849.5	23.8	666	9	US-09-331-631A-1	Sequence 1, Appl
31	849.5	23.8	666	14	US-10-147-095-1	Sequence 1, Appl
32	828	23.2	316	15	US-10-425-114-43905	Sequence 43905, A
33	828	23.2	316	15	US-10-425-114-49262	Sequence 49262, A
34	827	23.2	315	15	US-10-425-114-43971	Sequence 43971, A
35	816	22.9	301	15	US-10-425-114-49373	Sequence 49373, A
36	791	22.2	590	9	US-09-331-631A-8	Sequence 8, Appl
37	791	22.2	590	14	US-10-147-095-8	Sequence 8, Appl
38	790	22.2	300	15	US-10-425-114-43597	Sequence 43597, A
39	788	22.1	296	15	US-10-425-114-43984	Sequence 43984, A
40	770	21.6	291	15	US-10-425-114-51411	Sequence 51411, A
41	761	21.4	149	15	US-10-100-303A-54	Sequence 54, Appl
42	737.5	20.7	525	9	US-09-331-631A-7	Sequence 7, Appl
43	737.5	20.7	525	14	US-10-147-095-7	Sequence 7, Appl
44	732	20.6	276	15	US-10-425-114-44106	Sequence 44106, A
45	686	19.3	266	15	US-10-425-114-42946	Sequence 42946, A

ALIGNMENTS

RESULT 1

US-09-847-208-28
; Sequence 28, Application US/09847208
; Publication No. US20030082190A1
; GENERAL INFORMATION:
; APPLICANT: Saxon, Andrew
; APPLICANT: Zhang, Ke
; APPLICANT: Zhu, Daocheng
; TITLE OF INVENTION: FUSION MOLECULES AND TREATMENT OF
; TITLE OF INVENTION: IGE-MEDIATED ALLERGIC DISEASES
; FILE REFERENCE: UC67.002A
; CURRENT APPLICATION NUMBER: US/09/847,208
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 177
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; TYPE: PRT
; ORGANISM: Arachis hypogaea (peanut)
US-09-847-208-28

Alignment Scores:

Pred. No.: 2,03e-261 Length: 626
Score: 3286.00 Matches: 626
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 92.25% Indels: 0
DB: 10 Gaps: 0

US-10-728-051-1 (1-2032) x US-09-847-208-28 (1-626)

QY 50 ATGAGAGGAGGTTCTCCTACGTATGCTGCTAGGAGATCCTTCTCGCTTCAGTT 109
DB 1 MetArgGlyArgValSerProLeuMetLeuLeuGlyLeuValLeuAlaSerVal 20
QY 110 TCTGCAACGATCCAGTATCATCCTTACCAGAAGAAAACAGAACCCCTGCGCCAG 169
DB 21 SerAlaThrHisAlaLysSerProTyrGlnLysLysThrGluAsnProCysAlaGln 40
QY 170 AGTGCTCCAGAGTTGTCAACAGGAACCGGATGACTTGAAGCAAAAGCATCGAGTCT 229
DB 41 ArgCysLeuGlnSerCysGlnGlnGluProAspAspLeuLysGlnLysAlaCysGluSer 60
QY 230 CGTGCAACGATCGAGTATGATCCTCGTTGTCTATGATCCTCGAGGACACTGCG 289
DB 61 ArgCysThrLysLeuGluTyrAspProArgCysValTyrAspProArgGlyHisThrGly 80
QY 290 ACCACCAACCAAGTTCCCTCCAGGGAGCGACACAGTGGCGCCCAACCCCGGAGACTAC 349
DB 81 ThrThrAsnGlnArgSerProProGlyGluArgThrArgGlyArgGlnProGlyAspTyr 100
QY 350 GATGATCACCGCTCAACCCCAAGAGAGAGAGGAGCGGATGGGAGCCAGCTGAGCG 409
DB 101 AspAspAspArgArgGlnProArgGluGluGlyArgTyrGlyProAlaGlyPro 120
QY 410 AGGAGCGTCAAGAGAGAGAGACTGAGACACCAAGAGAGAGTTCGAGGCGCACCAAGT 469
DB 121 ArgGluArgGluArgGluGluAspTyrArgGlnProArgGluAspTyrArgGProSer 140
QY 470 CATCAGCAGCACCGAAATAAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 529
DB 141 HisGlnProArgLysIleArgProGluGlyArgGluGluGlnGluTyrGlyThr 160
QY 530 CCAGGTAGCATGTGAGGAGAGAAACATCTCGGAGAACACCTTCTACTTCCCGTCAAG 589
DB 161 ProGlySerHisValArgGluGluThrSerArgAsnAsnProPheTyrPheProSerArg 180
QY 590 CGTTTAGCACCCCTACGGGAACCAAAACGGTAGGATCCGGTCTCGCAGAGGTTGAC 649
DB 181 ArgPheSerThrArgTyrGlyAsnGlnAsnGlyArgIleArgValLeuGlnArgPheAsp 200
QY 650 CAAAGGTCAGGCAAGTTTCAAGATCTCAGAAATCACCGTATTGTGAGATCGAGGCCAAA 709
DB 201 GlnArgSerArgGlnPheGlnAsnLeuGlnAsnHisArgIleValGlnIleGluAlaLys 220
QY 710 CCTAACACTTGTCTTCCCAAGCAGCTGATGCTGATTAACATCTTGTATCCAGCAA 769
DB 221 ProAsnThrLeuValLeuProLysHisAlaAspAlaAspAsnIleLeuValIleGlnGln 240
QY 770 GGGCAAGCCCGTACCGTAGCAATGGCAATTAACAGAAAGAGCTTTAATCTTGACGAG 829
DB 241 GlyGlnAlaThrValThrValAlaAsnGlyAsnAsnArgLysSerPheAsnLeuAspGlu 260
QY 830 GGCCATGCACTCAGAAATCCCATCCGGTTTCAATTCCTTACATCTTGAACCGCCATGACAAC 889
DB 261 GlyHisAlaLeuArgIleProSerGlyPheIleSerTyrIleLeuAsnArgHisAspAsn 280
QY 890 CAGAACCTCAGAGTAGTAAATCTCCATGCGCTTAACACACCGCCGCTTTCAGGAT 949
DB 281 GlnAsnLeuArgValAlaLysIleSerMetProValAsnThrProGlyGlnPheGluAsp 300
QY 950 TTCTCCCGGAGCAGCGAGAGCAATCATCTACTTTCAGAGGCTTCAGCAGGAATACG 1009
DB 301 PhePheProAlaSerSerArgAspGlnSerSerTyrLeuGlnGlyPheSerArgAsnThr 320

QY 1010 TTGAGGCGCCCTTCAATGCGGAATTTCAATGAGATACGAGGGTCTGTTAGAGAGAAAT 1069
DB 321 LeuGluAlaAlaPheAsnAlaGluPheAsnGluIleArgArgValLeuLeuGluGluAsn 340
QY 1070 GCAGGAGTGAACAAGAGGAGAGGAGGAGGAGGAGGAGTACTCCGAGTAGGAGAAC 1129
DB 341 AlaGlyGlyGluGlnGluArgGlyGlnArgArgTyrSerThrArgSerSerGluAsn 360
QY 1130 AATGAAGAGTGAATGCTCAAAAGTGTCAAAAGGAGCAGCTTGAAGAACTTACTTAAGCAGCT 1189
DB 361 AsnGluGlyValIleValLysValSerLysGluHisValGluGluLeuThrLysHisAla 380
QY 1190 AAATCCGTCTCAAAAGAGGCTCCGAGAGAGAGAGAGATATCACCACCAATCAACTTG 1249
DB 381 LysSerValSerLysLysLysGlySerGluGluGluGlyAspIleThrAsnProIleAsnLeu 400
QY 1250 AGAGNAGCGAGCCCGATCTTCTAAACAATTCGGAGGTATTGTAGGTCAAGCCAGAC 1309
DB 401 ArgGluGlyGluProAspLeuSerAsnAsnPheGlyLysLeuPheGluValLysProAsp 420
QY 1310 AAGAAAGAACCCCGACTTCAGGACCTGACATGATGTCTCACTGTGTAGAGATCAAGAA 1369
DB 421 LysLysAsnProGlnLeuGlnAspLeuAspMetMetLeuThrCysValGluIleLysGlu 440
QY 1370 GGAGCTTTGATGCTCCCACTTCACTCAAGGCCATGTTTCTGCTCGTCAACAAA 1429
DB 441 GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValValAsnLys 460
QY 1430 GGAATCTGAAACCTTGAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1489
DB 461 GlyThrGlyAsnLeuGluLeuValAlaValArgLysGluGlnGlnArgGlyArgArg 480
QY 1490 GAAAGAGAGGAGGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1549
DB 481 GluGluGluGluAspGluGluGluGluGluGluGluGluGluGluGluValArgTyr 500
QY 1550 ACAGGAGGTTCAAGAGAGGAGGATGTGTATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1609
DB 501 ThrAlaArgLeuLysGluGlyAspValPheIleMetProAlaAlaHisProValAlaIle 520
QY 1610 AACGCTTCTCCGAACTCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1669
DB 521 AsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsnHisArg 540
QY 1670 ATCTTCTCTGAGGTGATAAGGACAAATGTATGATAGACACAGATAGAGAGAGAGAGAGAT 1729
DB 541 IlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGlnAlaLysAsp 560
QY 1730 TTAGCAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1789
DB 561 LeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGlnLysGluSer 580
QY 1790 CACTTTGTGAGTCTGCTCCTCAATCTCAATCTCAATCTCCTGCTGCTGCTGCTGCTGCTGCTGCT 1849
DB 581 HisPheValSerAlaArgProGlnSerGlnSerGlnSerProSerSerProGluLysGlu 600
QY 1850 TCTCTGAGAAAGAGATCAAGAGAGGAGAAACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1909
DB 601 SerProGluLysGluAspGlnGluGluAsnGlnGlyGlyLysGlyProLeuLeuSer 620
QY 1910 ATTTTGAAGGCTTTTAAAC 1927
DB 621 IleLeuLysAlaPheAsn 626

RESULT 2

US-10-728-806-2
; Sequence 2, Application US/10228806
; Publication No. US20030049237A1
; GENERAL INFORMATION:
; APPLICANT: Bannan, et al.
; TITLE OF INVENTION: Methods and Reagents for Decreasing Clinical Reactions
; TITLE OF INVENTION: to Allergy
; FILE REFERENCE: 2002834-0043

; CURRENT APPLICATION NUMBER: US/10/228,806
; CURRENT FILING DATE: 2002-08-26
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 626
; TYPE: PRT
; ORGANISM: Arachis hypogaea
US-10-228-806-2

Alignment Scores:
Pred. No.: 2,03e-261 Length: 626
Score: 3286.00 Matches: 626
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 92.25% Indels: 0
DB: 14 Gaps: 0

US-10-728-051-1 (1-2032) x US-10-228-806-2 (1-626)

Qy	50	ATGAGAGGGAGGGTTTCTCCACTGATGCTGTGCTAGGGATCCTTGTCCTGGCTTCAGTT	109
Db	1	MetArgGlyArgValSerProLeuMetLeuLeuLeuGlyLeuValLeuAlaSerVal	20
Qy	110	TCGCAACGCATGCCAAGTCATACCTTACAGAAAGAAACAGAGAACCCCTGGCCCGAG	169
Db	21	SerAlaThrHisAlaLysSerProTyrGlnLysLysThrGluAenProCysAlaGln	40
Qy	170	AGGTGCCTCCAGAGTCTCAACAGAAACCGATGACATTGAAGCAAAAGGATCGAGTCT	229
Db	41	ArgCysLeuGlnSerCysGlnGlnGlnProAspLeuLysGlnLysAlaCysGlnSer	60
Qy	230	CGCTGCACCAAGCTCGAGTATGATCCTCGTTGTGCTTATGATCCTCGAGGACACACTGGC	289
Db	61	ArgCysThrLysLeuGlnTyrAspProArgCysValTyrAspProArgGlyHisThrGly	80
Qy	290	ACACCAACCAAGTTCCTCCCTCAGGGAGCGGACAGTCGGCCGCCAACCCGGAGACTAC	349
Db	81	ThrThrAenGlnArgSerProProGlyGluArgThrArgGlyArgGlnProGlyAspTyr	100
Qy	350	GATGATCACCGCTCAACCCCGAAGAGAGAGAGCGCGATGGGGACACGCTGGAGCCG	409
Db	101	AspAspArgArgGlnProArgArgGluGluGluGluGluGluGluGluGluGluGluGlu	120
Qy	410	AGGAGCGGTGAAGAGAGAGAGTGGAGACAACCAAGAGAGAGATTTGGAGCGCACCAAGT	469
Db	121	ArgGluArgGluArgGluGluAspTyrArgGlnProArgGluAspTyrArgArgProSer	140
Qy	470	CATCAGACGCCAGAAATAAGCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	529
Db	141	HisGlnGlnProArgLysIleArgProGluGluGluGluGluGluGluGluGluGluGlu	160
Qy	530	CCAGGTAGCCATGTAGGAGAGAACATCTCGGAACACCTTCTACTTCCCGTCAAGG	589
Db	161	ProGlySerHisValArgGluGluThrSerArgAsnAsnProPheTyrPheProSerArg	180
Qy	590	CGGTTTAGCCACCCGCTACGGGAACCAAAACGGTAGGATCCGGGTCTCGACAGAGTTTGAC	649
Db	181	ArgPheSerThrArgTyrGlyAsnGlnAsnGlyArgIleArgValLeuGlnArgPheAsp	200
Qy	650	CAAGGTCAAGGCAGTTTCAGAAATCTCCAGATCACCGTATTGTGCGAGATCGAGGCCAAA	709
Db	201	GlnArgSerArgGlnPheGlnAsnLeuGlnAsnHisArgIleValGlnIleGluAlaLys	220
Qy	710	CCTAACACTCTGTCTTCCCAACGACGCTGATCTGATACATCTTGTATCCAGCAA	769
Db	221	ProAsnThrLeuValLeuProLysHisAlaAspAlaAspAsnIleLeuValIleGlnGln	240
Qy	770	GGGCAAGCCACCGTAGCAAAATGGCAATTAACAGAAAGAGCTTTAATCTTTGACGAG	829
Db	241	GlyGlnAlaThrValThrValAlaAsnGlnAsnArgLysSerPheAsnLeuAspGlu	260
Qy	830	GGCCATGCACTCAGAAATCCCATCCGGTTTCATTTCTTACATCTTTGAACCCGCCATGACAAC	889

Db	261	GlyHisAlaLeuArgIleProSerGlyPheIleSerTyrIleLeuAenArgHisAspAsn	280
Qy	890	CAGAACTCAGAGTAGCTAAATCTCCATCCCGTTTAAACACACCCGCCAGTTTGAGGAT	949
Db	281	GlnAsnLeuArgValAlaLysIleSerMetProValAsnThrProGlyGlnPheGluAsp	300
Qy	950	TTCTTCCCGCAGCAGCCGAGACCAATCATCTTCTTTCAGGGCTTTCAGCAGAAATACG	1009
Db	301	PhePheProAlaSerSerArgAspGlnSerSerTyrLeuGlnGlyPheSerArgAsnThr	320
Qy	1010	TTGGAGCCCGCTTCAATCGGAATTCATGAGATACGGAGGCTGCTGTAGAGACAAT	1069
Db	321	LeuGluAlaAlaPheAenAlaGluPheAsnGluIleArgArgValLeuLeuGluGluAsn	340
Qy	1070	GCAGGAGGTGAGCAAGAGGAGAGAGCGAGCGATGGAGTACTCCGAGTAGTGAGAAC	1129
Db	341	AlaGlyGlyGluGlnGluArgGlyGlnArgTyrSerThrArgSerSerGluAsn	360
Qy	1130	AATGAAGGAGTATAGTCAAAAGTGTCAAAGGACAGCTTGAAGAACTTACTAAGACGCT	1189
Db	361	AsnGluGlyValIleValLysValSerLysGluHisValGluGluLeuThrLysHisAla	380
Qy	1190	AAATCCGTCTCAAGAAAGCTCCGAGAGAGGAGATATCACCACCCCAATCAACTTG	1249
Db	381	LysSerValSerLysLysGlySerGluGluGluGlyAspIleThrAsnProIleAsnLeu	400
Qy	1250	AGAGAAGCGAGCCGATCTTTCTAACTTTGGGAAGTTATTGAGGTGAAGCCAGAC	1309
Db	401	ArgGluGlyGluProAspLeuSerAsnAsnPheGlyLysLeuPheGluValLysProAsp	420
Qy	1310	ARGAAGAACCCCGACTTACAGGACCTGGACATGATGATCTACCTGTGTAGAGATCAAA	1369
Db	421	LysLysAsnProGlnLeuGlnAspLeuAspMetMetLeuThrCysValGluIleLysGlu	440
Qy	1370	GGAGCTTGTGATGCTCCACACTTCAACTCAAGCCCATGCTTATCTCGCTCGTCAACAA	1429
Db	441	GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValAsnLys	460
Qy	1430	GGAACTGGAACTTGAACCTCGTGGCTGTAAAGAAAGAGCAACAAAGAGGGAGCGCG	1489
Db	461	GlyThrGlyAsnLeuGluLeuValAlaValArgLysGluGlnGlnArgArg	480
Qy	1490	GAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1549
Db	481	GluGluGluGluAspGluAspGluGluGluGluGluGluGluGluGluGluGluGluGlu	500
Qy	1550	ACAGCGAGGTGAAGNAGCGGATGTTTCATCATGCCAGCAGCTCATCCAGTAGCCATC	1609
Db	501	ThrAlaArgLeuLysGluGlyAspValPheIleMetProAlaAlaHisProValAlaIle	520
Qy	1610	AACGCTTCTCCGAACCTCCATCTGCTTGGCTTCGGTATCAACGCTGAAAAACAACAG	1669
Db	521	AsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsnAsnHisArg	540
Qy	1670	ATCTTCTTCAGGTGATAAGCAATGTGATAGACAGATAGAGAGAGAGAGAGAGAGAT	1729
Db	541	IlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGlnAlaLysAsp	560
Qy	1730	TTAGCATTCCTCGGTCCGGTGAACAGTTGAGAGCTCATCAAAACCCAGAGAGATCT	1789
Db	561	LeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGlnLysGluSer	580
Qy	1790	CACCTTGTGAGTGTCTCTCTCAATCTCAATCTCAATCTCCGTCTCCGTCTCCGTGAGAA	1849
Db	581	HisPheValSerAlaArgProGlnSerGlnSerProSerSerProGluLysGlu	600
Qy	1850	TCTCCTGAGAAAGAGATCAAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1909
Db	601	SerProGluLysGluAspGlnGluGluGluGluGluGluGluGluGluGluGluGluGlu	620
Qy	1910	ATTTTGAAGCTTTTAAAC	1927

Db 621 IleLeuLysAlaPheAsn 626

RESULT 3
US-10-100-303A-7
; Sequence 7, Application US/10100303A
; Publication No. US20030202980A1
; GENERAL INFORMATION:
; APPLICANT: Caplan, et al.
; TITLE OF INVENTION: Methods and Reagents for Decreasing Clinical Reaction
; TITLE OF INVENTION: to Allergy
; FILE REFERENCE: 2002834-0166
; CURRENT APPLICATION NUMBER: US/10/100,303A
; CURRENT FILING DATE: 2002-03-18
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; TYPE: PRT
; ORGANISM: Arachis hypogaea, Prot/Nucleo-Ara h 1
US-10-100-303A-7

Alignment Scores:
Pred. No.: 2,03e-261 Length: 626
Score: 3286.00 Matches: 626
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 92.25% Indels: 0
DB: 15 Gaps: 0

US-10-728-051-1 (1-2032) x US-10-100-303A-7 (1-626)

QY 50 ATGAGAGGAGGGTTCTCCACTGATGCTGCTAGGAGTCTTGTCTCGGCTTCAGTT 109
Db 1 MetArgGlyArgValSerProLeuMetLeuLeuLeuGlyIleLeuValLeuAlaSerVal 20

QY 110 TCTGCAACGATGCCAAGTATCATCACCCTTACAGAGAAACAGAGAACCCCTCGGCCAG 169
Db 21 SerAlaThrHisAlaLysSerSerProTyrGlnLysLysThrGluAsnProCysAlaGln 40

QY 170 AGTGCTCCAGAGTCTCAACAGGACCGGATGCTTGAAGCAAAAGGATCCGAGTCT 229
Db 41 ArgCysLeuGlnSerCysGlnGlnGlnProAspAspLeuLysGlnLysAlaCysGluSer 60

QY 230 CGCTGCACCAAGCTCGAGTATGATCTCTGTTGTCTATGATCCTCGAGGACACTGGC 289
Db 61 ArgCysThrLysLeuGluTyrAspProArgCysValTyrAspProArgGlyHisThrGly 80

QY 290 ACCACCAACCACTTCCCTCCAGGAGGGGACAGTGGCCGCCCAACCCGGAGCTAC 349
Db 81 ThrThrAsnGlnArgSerProProGlyGluArgThrArgGlyArgGlnProGlyAspTyr 100

QY 350 GATGATGACCGCTCAACCCGAGAGAGAGAGGAGGCGGATGGGACCAAGCTGGACCG 409
Db 101 AspAspAspArgArgGlnProArgArgGluGluGlyArgTyrProAlaGlyPro 120

QY 410 AGGAGCGTGAAGAGAGAGAGTGGAGACACCAAGAGAGATTCGAGCGGACCAAGT 469
Db 121 ArgGluArgGluArgGluGluAspTyrArgGlnProArgGluAspTyrArgArgProSer 140

QY 470 CATCAGCAGCCAGGAAATAGCCCGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 529
Db 141 HisGlnGlnProArgLysIleArgProGluGlyArgGluGlyGluGlnGlnTyrPglyThr 160

QY 530 CCAGTATGACCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 589
Db 161 ProGlySerHisValArgGluGluThrSerArgAsnAsnProPheTyrPheProSerArg 180

QY 590 CGGTTTACGACCCGCTACGGAAACCAAAACCGTAGGATCCGGGTCCTGCAGAGGTTTAC 649
Db 181 ArgPheSerThrArgTyrGlyAsnGlnAsnGlyArgIleArgValLeuGlnArgPheAsp 200

QY 650 CAAGGTCAGGCGAGTTTCAGAACTCCAGAAATCACCGTATTGTGCAGATCGAGGCCAAA 709
Db 1 MetArgGlyArgValSerProLeuMetLeuLeuLeuGlyIleLeuValLeuAlaSerVal 20

201 GlnArgSerArgGlnPheGlnAsnLeuGlnAsnHisArgIleValGlnIleGluAlaLys 220
710 CCTAACACTCTTGTCTTCCCAAGCAGCTGATGCTGATACATCTTGTATTATCAGCAA 769
221 ProAsnThrLeuValLeuProLysHisAlaAspAlaAspAsnIleLeuValIleGlnGln 240
770 GGGCAAGCCACCGTACCGTAGCAAAATGGCAATTAACAGAAAGAGCTTTAATCTTGACGAG 829
241 GlyGlnAlaThrValThrValAlaAsnGlyAsnAsnArgLysSerPheAsnLeuAspGlu 260
830 GGCCATGCACTCAGAAATCCCATCCGTTTCAATTCCTACATCTTGAACCGCATGACAAAC 889
261 GlyHisAlaLeuArgIleProSerGlyPheIleSerTyrIleLeuAsnArgHisAspAsn 280
890 CAGAACCTCAGAGTAGCTAAATCTCCATGCGCTTAACACACCCGGCAGCTTTGAGAT 949
281 GlnAsnLeuArgValAlaLysIleSerMetProValAsnThrProGlyGlnPheGluAsp 300
950 TTCTTCCCGGAGCAGCCGAGACCAATCATCTTCTGAGGGCTTCAGCAGGATACG 1009
301 PhePheProAlaSerArgAspGlnSerSerTyrLeuGlnGlyPheSerArgAsnThr 320
1010 TTGGAGGCGCTTCAATGCCGAATTCATGAGATACGAGGCTGCTGTAGAACAGAT 1069
321 LeuGluAlaAlaPheAsnAlaGluPheAsnGluIleArgValLeuLeuGluGluAsn 340
1070 GCAGGAGGTGAGCAAGAGAGAGAGGAGGAGGAGGAGTACTCGAGTAGTAGAGAAC 1129
341 AlaGlyGlyGluGlnGluGluArgGlyGlnArgArgTyrSerThrArgSerSerGluAsn 360
1130 AATGAGAGGTATAGTCAAGTGTCAAAGAGGACGCTTGAAGACTTACTTAAGCACCT 1189
361 AsnGluGlyValIleValLysValSerLysGluHisValGluGluLeuThrLysHisAla 380
1190 AATCCGCTCAAGAAAGGCTCCGAGAGAGGAGATATCACCACCCCAATCACTTG 1249
381 LysSerValSerLysLysGlySerGluGluGluGlyAspIleThrAsnProIleAsnLeu 400
1250 AGAGAGGCGAGCCCGATCTTTCTAACAACTTTGGAGAGTTATTTGAGGTGAAGCCAGAC 1309
401 ArgGluGlyGluProAspLeuSerAsnAsnPheGlyLysLeuPheGluValLysProAsp 420
1310 AGAGAGACCCCGACCTTCAGGACCTGGACATGATGCTCACCTGTGTAGAGATCAAGAA 1369
421 LysLysAsnProGlnLeuGlnAspLeuAspMetLeuThrCysValGluLysGlu 440
1370 GGAGCTTTTATGCTCCCACTTCAACTCAAGGCGATGTTTATCTGCTGCTCAACAAA 1429
441 GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValAsnLys 460
1430 GGAATCGGAACTTGAATCTGCTGCTGTAAGAAAGAGAGAGAGAGAGAGAGAGAGAG 1489
461 GlyThrGlyAsnLeuGluValAlaValArgLysGluGlnGlnArgGlyArgArg 480
1490 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1549
481 GluGluGluGluAspGluAspGluGluGluGluGluGluGluGluGluGluGluGlu 500
1550 ACAGCGAGGTGTAAGAGAGCGATGTTTCATCATGCCAGCAGCTCATCAGTAGGCATC 1609
501 ThrAlaArgLeuLysGluGlyAspValPheIleMetProAlaAlaHisProValAlaIle 520
1610 AACGCTTCTCCGAATCTCATCTGCTTGGTTCGTTATCAACGCTGAAACCAACACAGA 1669
521 AsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsnAsnHisArg 540
1670 ATCTTCTTCTGAGGTGATAGGACATGTGATAGACAGATAGAGAGAGAGAGAGAGAT 1729
541 IlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGlnAlaLysAsp 560
1730 TTAGCATTCCTCGGTGCGGTGAACAGTTGAGAGAGCTCATCAAAACAGAGAGATCT 1789
561 LeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGlnLysGluSer 580

```
QY 1790 CACTTGTGAGTCTCTCTCAATCTCAATCTCAATCTCCGTCGTCTCTCGAAGAG 1849
D 581 HisPheValSerAlaArgProGlnSerGlnSerGlnSerProSerProGluLysGlu 600
QY 1850 TCTCTCAGAAAGAGGATCAAGAGGAGAAACCAAGGAGGAGGTCCTCACTCTTCA 1909
D 601 SerProGlnLysGluAspGlnGluGluGluAsnGlnGlyLysGlyProLeuLeuSer 620
QY 1910 ATTTGAAGGCTTTTAAAC 1927
D 621 IleLeuLysAlaPheAsn 626

RESULT 4
US-10-245-871-10
; Sequence 10, Application US/10245871
; Publication No. US20030235594A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT
; TITLE OF INVENTION: II-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: REH-2013
; CURRENT APPLICATION NUMBER: US/10/245,871
; CURRENT FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: 10/197,000
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 905
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 626
; TYPE: PRT
; ORGANISM: Arachis hypogaea
US-10-245-871-10

Alignment Scores:
Pred. No.: 2,03e-261 Length: 626
Score: 3286.00 Matches: 626
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 92.25% Indels: 0
DB: 15 Gaps: 0

US-10-728-051-1 (1-2032) x US-10-245-871-10 (1-626)
QY 50 ATGAGAGGAGGTTTCTCTCACTGATGCTGTCTGTTAGGATCCTGTCTCGCTTCAGTT 109
D 1 MetArgGlyArgValSerProLeuMetLeuLeuLeuGlyIleLeuValLeuAlaSerVal 20
QY 110 TCTGCAACGATGCCAAGTCAATCACCTTACCAGAAGAAACAGAGAACCCCTCGGCCAG 169
D 21 SerAlaThrHisAlaLysSerSerProTyrGlnLysLysThrGluAsnProCysAlaGln 40
QY 170 AGTGCTCCAGAGTTGTCAACAGGAACCGGATGACTTGAAGCAAAAGGCATCGAGTCT 229
D 41 ArgCysLeuGlnSerCysGlnGlnGluProAspAspLeuLysGlnLysAlaCysGluSer 60
QY 230 CGCTGCACCAAGCTCGAGTATGATCCTCGTTGTCTCTATGATCCTCGAGGACACTGGC 289
D 61 ArgCysThrLysLeuGluTyrAspProArgCysValTyrAspProArgGlyHisThrGly 80
QY 290 ACCACCAACCAAGCTTCCCTCCAGGGAGCGGACACGTCGGCCGCCAACCCCGAGACTAC 349
D 81 ThrThrAsnGlnArgSerProProGlyGluArgThrArgGlyArgGlnProGlyAspTyr 100
QY 350 GATGATGACCCGCTCAACCCGAGAGAGAGAGGAGCGATGGGACCCAGCTGGACCG 409
D 101 AspAspAspArgArgGlnProArgArgGluGluGlyGlyArgTrpGlyProAlaGlyPro 120
QY 410 AGGAGCGCTCAAGAGAGAGACTCGAGACACCAAGAGAGAGATTTGGAGCCCAAGT 469
D 121 ArgGluArgGluArgGluGluAspTrpArgGlnProArgGluAspTrpArgArgProSer 140
```

```
QY 470 CATCAGCAGCCACGAAAAATAAGCCCGAAGAGAGAGAGAGAACAAAGAGTGGGAACA 529
D 141 HisGlnGlnProArgLysIleArgProGluGlyArgGluGlyGluGlnGluTrpGlyThr 160
QY 530 CAGGTAGCCATGTGAGGGAAGAAACATCTCGGAACAACCCCTTCTACTTCCCGTCAAGG 589
D 161 ProGlySerHisValArgGluThrSerArgAsnAsnProPheTyrPheProSerArg 180
QY 590 CGGTTTAGCACCCGCTACGGGAACCAAAACGGTAGGATCCGGTCTCGACAGAGTTTGAC 649
D 181 ArgPheSerThrArgTyrGlyAsnGlnAsnGlyArgIleArgValLeuGlnArgPheAsp 200
QY 650 CAAAGGTCAGGCAGTTTCAGAAATCCAGAAATCACCGTATTGTGCAGATCGAGGCCAAA 709
D 201 GlnArgSerArgGlnPheGlnAsnLeuGlnAsnHisArgIleValGlnIleGluAlaLys 220
QY 710 CTTAACACTCTTGTCTTCTCCCAAGCAGCTGATGCTGATTAACATCTTGTATTCCAGCAA 769
D 221 ProAsnThrLeuValLeuProLysHisAlaAspAlaAspAsnIleLeuValIleGlnGln 240
QY 770 GGGCAAGCCACCGTAGCCGTAGCAAAATGGCAATAACAGAAAGAGCTTTAATCTTGACGAG 829
D 241 GlyGlnAlaThrValThrValAlaAsnGlyAsnAsnArgLysSerPheAsnLeuAspGlu 260
QY 830 GGCCATGCACCTCAGAAATCCCATCCGGTTCATTCTCTACATCTTGAACCCGCTATGACAAC 889
D 261 GlyHisAlaLeuArgIleProSerGlyPheIleSerTyrIleLeuAsnArgHisAspAsn 280
QY 890 CAGAACTCAGAGTAGCTAAATCTCCATGCCCGTTAAACACACCCCGGCAGTTTGAGGAT 949
D 281 GlnAsnLeuArgValAlaLysIleSerMetProValAsnThrProGlyGlnPheGluAsp 300
QY 950 TTCTTCCCGCGCAGCAGCAGCAGCAATCATCTCTCTGTCAGGGCTTCAGCAGCAATACG 1009
D 301 PhePheProAlaSerSerArgAspGlnSerSerTyrLeuGlnGlyPheSerArgAsnThr 320
QY 1010 TTGGAGCCCGCTTCAATCGGAATTCATAGATACGAGGAGGTGCTGTGTAGAGAGAAAT 1069
D 321 LeuGluAlaAlaPheAsnAlaGluPheAsnGluIleArgArgValLeuLeuGluAsn 340
QY 1070 GCAGGAGGTGAGCAAGAGCAGAGAGAGCGGCGATGAGTACTCGAGTAGTAGAAC 1129
D 341 AlaGlyGlyGluGlnGluArgGlyGlnArgTrpSerThrArgSerSerGluAsn 360
QY 1130 AATGAAGGAGTGTAGTCAAAAGTGTCAAAGGACGACGTTGAAGAACTTACTTAAGACGCT 1189
D 361 AsnGluGlyValIleValLysValSerLysGluHisValGluGluLeuThrLysHisAla 380
QY 1190 AATCCGCTCTCAAAGAAAGGCTCCGAAGAGAGGAGAGATATCACCAACCCCAATCAACTTG 1249
D 381 LysSerValSerLysLysGlySerGluGluGlyAspIleThrAsnProIleAsnLeu 400
QY 1250 AGAAGAGGCGAGCCGATCTTTCTAACACTTTGGGAAGTTATTTGAGGTGAAGCCAGAC 1309
D 401 ArgGluGlyGluProAspLeuSerAsnAsnPheGlyLysLeuPheGluValLysProAsp 420
QY 1310 AAGAAGAACCCCGCAGCTTCAGGACCTGACATGCTCACCTGTGTAGAGATCAAGAA 1369
D 421 LysLysAsnProGlnLeuGlnAspLeuAspMetLeuThrCysValGluIleLysGlu 440
QY 1370 GGAGCTTTTGTATGTCCTCCACACTTCAACTCAAAGGCCATGGTTATCGTCGTCGCAACAAA 1429
D 441 GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValValAsnLys 460
QY 1430 GGAACCTGGAAACCTTGAACCTGTCGCTGTAAGAAAAGAGCAACACAGAGGGGACGGCGG 1489
D 461 GlyThrGlyAsnLeuGluLeuValAlaValArgLysGluGlnGlnArgGlyArgArg 480
QY 1490 GAAGAGAGGAGGAGCAACCAAGAGAGAGGAGGAGGAGTAACAGAGAGGTCGCTAGGTAC 1549
D 481 GluGluGluGluAspGluAspGluGluGluGluGlySerAsnArgGluValArgArgTyr 500
```

```
QY 1550 ACAGCAGGCTGAAGACGCGATGTTCATCATGCGCAGCAGCTCATCCAGTAGCCATC 1609
Db 501 ThrAlaArgLeuLysGluGlyAspValPheIleMetProAlaAlaHisProValAlaIle 520
QY 1610 AACGCTTCTCCGAACTCCATCTGCTTGGCTTCGGTATCAACGCTGAAACACACAGA 1669
Db 521 AsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsnHisArg 540
QY 1670 ATCTTCTTCAGGTGATAGGACATGATAGACCATAGACCATAGACCAAGCGAGGAT 1729
Db 541 IlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGlnAlaLysAsp 560
QY 1730 TTAGCATTCCTCGGTCGGTGAAACAAGTTGAGAAGCTCATCAAAACACAGAAAGAAATCT 1789
Db 561 LeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGlnLysGluSer 580
QY 1790 CACTTTGTAGTCTCGTCTCAATCTCAATCTCAATCTCGGTCGTCCTCAGAAAGAG 1849
Db 581 HisPheValSerAlaArgProGlnSerGlnSerGlnSerProSerProGluLysGlu 600
QY 1850 TCTCCTGAGAAAGAGGATCAAGAGGAGGAAACCAAGGAGGAGGAGGAGGAGGAGGAGG 1909
Db 601 SerProGluLysGluAspGlnGluGluGluAsnGlnGlyLysGlyProLeuLeuSer 620
QY 1910 ATTTTGAAGGCTTTTAAAC 1927
Db 621 IleLeuLysAlaPheAsn 626

RESULT 5
US-10-253-286-10
; Sequence 10, Application US/10253286
; Publication No. US20040058881A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT
; TITLE OF INVENTION: 11-KBY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: REH-2015
; CURRENT APPLICATION NUMBER: US/10/253,286
; CURRENT FILING DATE: 2003-01-13
; PRIOR APPLICATION NUMBER: 10/197,000
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 905
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 10
; LENGTH: 626
; TYPE: PRT
; ORGANISM: Arachis hypogaea
US-10-253-286-10

Alignment Scores:
Pred. No.: 2,03e-261 Length: 626
Score: 3286.00 Matches: 626
Percent Similarity: 100.00% Conservaive: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 92.25% Indels: 0
DB: 15 Gaps: 0

US-10-728-051-1 (1-2032) x US-10-253-286-10 (1-626)
QY 50 ATGAGAGGAGGGTTTCTCCACTGATGCTGTGTCTAGGATCCTTGTCTGGCTTCAGTT 109
Db 1 MetArgGlyArgValSerProLeuMetLeuLeuLeuGlyIleLeuValLeuAlaSerVal 20
QY 110 TCTGCAACCCATGCCAAGTCATCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 169
Db 21 SerAlaThrHisAlaLysSerSerProTyrGlnLysLysThrGluAsnProCysAlaGln 40
QY 170 AGTGTCTCCAGAGTTGTCAACAGAGAACCGGATGATCTTGAAGCAAAAGCGATGCGAGTCT 229
Db 41 ArgCysLeuGlnSerCysGlnGlnGluProAspAspLeuLysGlnLysAlaCysGluSer 60
```

```
QY 230 CGTGTCCACCAAGCTCGAGTATGATCTCGTGTGTCTATGATCTCTCGAGGACACACTGGC 289
Db 61 ArgCysThrLysLeuGluTyrAspProArgCysValTyrAspProArgGlyHisThrGly 80
QY 290 ACCACCAACCAAGTTCCTCCAGGGAGCGACACGTCGGCGCCCAACCCCGAGACTAC 349
Db 81 ThrThrAsnGlnArgSerProGlyGluArgThrArgGlyArgGlnProGlyAspTyr 100
QY 350 GATCATCACCCGCTCAACCCCGAAGAGAGAAAGGAGCGATGGGACACAGCTGGACCG 409
Db 101 AspAspAspArgArgGlnProArgArgGluGluGlyArgTrpGlyProAlaGlyPro 120
QY 410 AGGAGCGGTGAAGAGAGAAAGACTGAGACACACCAAGAGAAAGATTGGAGCGCACCAAGT 469
Db 121 ArgGluArgGluArgGluAspTrpArgGlnProArgGluAspTrpArgArgProSer 140
QY 470 CATCAGCAGCCACGAAATAAGGCCCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 529
Db 141 HisGlnGlnProArgLysIleArgProGluGlyArgGluGlyGluGlnGluTrpGlyThr 160
QY 530 CCAGGTAGCCATGTGAGGAGAAACAATCTCGGAAACAACCTTTCTACTTCCCGTCAAGG 589
Db 161 ProGlySerHisValArgGluThrSerArgAsnAsnProPheTyrPheProSerArg 180
QY 590 CGGTTTAGCACCCGCTACGGGAACCAAAACCGTAGGATCCGGGTCTCGCAGAGGTTTGAC 649
Db 181 ArgPheSerThrArgTyrGlyAsnGlnAsnGlyArgIleArgValLeuGlnArgPheAsp 200
QY 650 CAAAGGTCAAGCGAGTTTTCAGAATCTCCAGAATCACCGTATTTGTCAGATCGAGGCCAAA 709
Db 201 GlnArgSerArgGlnPheGlnAsnLeuGlnAsnHisArgIleValGlnIleGluAlaLys 220
QY 710 CCTAACACTCTTGTTCCTCCAGCAGCGTCATCTGATACATCATCTTGTATTCAGCAA 769
Db 221 ProAsnThrLeuValLeuProLysHisAlaAspAlaAspAsnIleLeuValIleGlnGln 240
QY 770 GGGCAAGCACCGTCACCGTAGCAAAATGGCAATACAGAAAGAGCTTTAATCTTGACGAG 829
Db 241 GlyGlnAlaThrValThrValAlaAsnGlyAsnAsnArgLysSerPheAsnLeuAspGlu 260
QY 830 GGCCATGCATCAGAAATCCCATCCGTTTCATTTCTTCTACATCTTGAACCCGCTACCAAC 889
Db 261 GlyHisAlaLeuArgIleProSerGlyPheIleSerTyrIleLeuAsnArgHisAspAsn 280
QY 890 CAGAACCTCAGAGTAGCTAAATCTCCATGCCGCTTAACACACCCGCGCAGTTTGAGGAT 949
Db 281 GlnAsnLeuArgValAlaLysIleSerMetProValAsnThrProGlyGlnPheGluAsp 300
QY 950 TTCTTCCCGCGCAGCAGCCAGACCAATCATCTTCTGAGGCGCTTCAGCGGATACG 1009
Db 301 PhePheProAlaSerSerArgAspGlnSerSerTyrLeuGlnGlyPheSerArgAsnThr 320
QY 1010 TTGGAGGCGCTTCATTCGGAATTCATGAGATACGGAGGGTCTGTTAGAAGAGAT 1069
Db 321 LeuGluAlaAlaPheAsnAlaGluPheAsnGluIleArgArgValLeuLeuGluAsn 340
QY 1070 GCAGGAGGTGAGCAAGAGAGAGAGCGCAGAGCGATGGAGTACTCGGAGTAGTGAGAAC 1129
Db 341 AlaGlyGlyGluGlnGluGluArgGlyGlnArgArgTrpSerThrArgSerSerGluAsn 360
QY 1130 AATGAAGAGTGTATGATCAAGTGTCAAAGGAGCAGTTGAAAGAACTTACTAAGCAGCGCT 1189
Db 361 AsnGluGlyValIleValLysValSerLysGluHisValGluGluLeuThrLysHisAla 380
QY 1190 AATTCGCTCTCAAAGAAAGCGCTCCGAAGAGAGGAGATATCACCACCAATCAACTTG 1249
Db 381 LysSerValSerLysLysGlySerGluGluGluGlyAspIleThrAsnProIleAsnLeu 400
QY 1250 AGAAGAGCGCAGCCCGATCTTTCTAAACAATTTGGAAAGATTATTGAGTGAGCCAGAC 1309
Db 401 ArgGluGlyGluProAspLeuSerAsnAsnPheGlyLysLeuPheGluValLysProAsp 420
QY 1310 AAGAAGAACCCCGCTTCAGGACCTGGACATGATCTCACCTGTCTAGAGATCAAGAA 1369
```


421	Db	 LySlyAsnProGlnLeuGlnAspLeuAspMetLeuThrCysValGluIleLeysGlu	440
1370	Qy	GGAGCTTTGATGCTCCACACTTCAACTCAAAAGGCCATGGTTATCGTCGTCGCAACAAA	1429
441	Db	 GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValValAsnLys	460
1430	Qy	GGAACTGGAAACCTTGAACTCGTGGCTGTGAAGAAAGACGACACACAGAGGGGACGGCGG	1489
461	Db	 GlyThrGlyAsnLeuGluLeuValAlaValArgLysGluGlnGlnArgGlyArgArg	480
1490	Qy	GAAGAAGAGGAGGACGAAGACGAAGAAGAGAGGAGGAGAAAGTAACAGAGAGGTGCGTAGGTAC	1549
481	Db	 GluGluGluGluAspGluAspGluGluGluGlySerAsnArgGluValArgArgTyr	500
1550	Qy	ACAGCGAGGTGAAGGAGGCGATGCTGTTTCATATGCCAGCAGCTCATCCAGTAGCCATC	1609
501	Db	 ThrAlaArgLeuLysGluGlyAspValPheIleMetProAlaIaHisProValIaIle	520
1610	Qy	AAGCGTTCTCCGAACCTCCACTCTGCTGGGCTTCGGTATCAACGCTGAAAAACAACACAGA	1669
521	Db	 AsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsnAsnHisArg	540
1670	Qy	ATCTTCTTCGAGGTGATAAGACAATGTGATAGACACAGATAGAGAAGCAGCGAAGGAT	1729
541	Db	 IlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGlnAlaLysAsp	560
1730	Qy	TTAGCATTCCTCGGTCCGGTGAACAAGTTGAGAAGCTCATCAAAAAACCAAGAGGAATCT	1789
561	Db	 LeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGlnLysGluSer	580
1790	Qy	CACTTTGAGTGCTCGTCTCCTCAATCTCAATCTCCGTCGTCTCCTGAGAAAGAG	1849
581	Db	 HisPheValSerAlaArgProGlnSerGlnSerGlnSerProSerSerProGluLysGlu	600
1850	Qy	TCTCTCTGAGAAAGAGGATCAAGAGGAGGAAAAACCAAGGAGGAGGCTCCACTCTCTTCA	1909
601	Db	 SerProGluLysGluAspGlnGluGluLysGlnGlyLysGlyProLeuLeuSer	620
1910	Qy	ATTTTGAAGCTTTTAAAC	1927
621	Db	 IleLeuLysAlaPheAsn	626

RESULT 6

```

US-10-899-551-2
; Sequence 2, Application US/10899551
; Publication No. US20050063994A1
; GENERAL INFORMATION:
; APPLICANT: Caplan, Michael J.
; APPLICANT: Burks, A. Wesley
; APPLICANT: Sampson, Hugh A.
; APPLICANT: Howard, Sosin B.
; APPLICANT: Bottomly, Kim H.
; TITLE OF INVENTION: Methods and Reagents for Decreasing Clinical Reaction to Allergy
; FILE REFERENCE: 2002834-0233
; CURRENT APPLICATION NUMBER: US/10/899,551
; CURRENT FILING DATE: 2004-07-26
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 626
; TYPE: PRT
; ORGANISM: species Arachis hypogaea
US-10-899-551-2

```

Alignment Scores:					
Pred. No.:	2.03e-261				
Score:	3286.00	Length:	626		
Percent Similarity:	100.00%	Matches:	626		
Best Local Similarity:	100.00%	Conservative:	0		
Query Match:	92.25%	Mismatches:	0		
DB:	17	Indels:	0		
		Gaps:	0		

QY	1130	AATGAAGGAGTGAAGTCAAAGTGTCAAAGGAGCAGCTTGAAGAACTTACTAAGCAGCT	1189
DB	361	AsnGluGlyValIleValLysValSerLysGluHisValGluGluLeuThrLysHisAla	380
QY	1190	AAATCCGTCTCAAGAAAGGCTCGAAGAAGAGGGAGATATCAACCAACCAACTCACTTG	1249
DB	381	LysSerValSerLysLysGlySerGluGluGluGlyAspIleThrAsnProIleAsnLeu	400
QY	1250	AGACAAAGCCGAGCCGAGCTTTCTTCAACAACTTTGGGAAGTTATTGTAGGTGAAGCCAGAC	1309
DB	401	ArgGluGlyGluProAspLeuSerAsnAsnPheGlyLysLeuPheGluValLysProAsp	420
QY	1310	AAGAAGAAACCCGAGCTTCAGGACCTGGACATGATGTCTACCTGTGTAGAGATCAAGAA	1369
DB	421	LysLysAsnProGlnLeuGlnAspLeuAspMetMetLeuThrCysValGluIleLysGlu	440
QY	1370	GGAGCTTTGATGCTCCCACTCTTCAACTCAAAGGCCATGGTTATCGTCGTCGTCACAAA	1429
DB	441	GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValValAsnLys	460
QY	1430	CGAACTGGAACCTTGAACTCGTGGCTGTAAAGAAAGAGCAACAAAGAGGGGAGCGCG	1489
DB	461	GlyThrGlyAsnLeuGluLeuValAlaValArgLysGluGlnGlnGlnArgGlyArgArg	480
QY	1490	GAAGAAGAGGAGGAGGAGAGAGAGAGAGAGGAGGAGAGTACACAGAGGTCGTCAGCTAC	1549
DB	481	GluGluGluGluAspGluAspGluGluGluGluGlySerAsnAsgluValArgArgTyr	500
QY	1550	ACAGCAGGTTGAAGGAAGGCGATGTGTTTCATCATGCCAGAGCTCATCCAGTAGCCATC	1609
DB	501	ThrAlaArgLeuLysGluGlyAspValPheIleMetProAlaIleHisProValAlaIle	520
QY	1610	AACGCTTCCTCGAACTCCATCTGCTTCGGTTCGGTATCAACGCTGAAACAAACACAGA	1669
DB	521	AsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsnAsnHisArg	540
QY	1670	ATCTTCCTTGCAAGTGATAAGGACAATGTGTATAGACCAGATAGAGAACGACGAGGAT	1729
DB	541	IlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGlnAlaLysAsp	560
QY	1730	TTAGCATTTCCCTGGGTCCGGTGAAACAGTTGAGAAGCTCATCAAAACACAGAAGGAATCT	1789
DB	561	LeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGlnLysGluSer	580
QY	1790	CACTTTGTGAGTCTCGTCTCTCAATCTCAATCTTCGCTGCTCTCTCTGAGAAAGAG	1849
DB	581	HisPheValSerAlaArgProGlnSerGlnSerGlnSerProSerSerProGluLysGlu	600
QY	1850	TCTCTCTGAAGAGAGGATCAAGAGAGGAGAAACCAAGGAGGAGGGTCCACTCTCTTCA	1909
DB	601	SerProGluLysGluAspGlnGluGluAsnGlnGlyGlyLysGlyProLeuLeuSer	620
QY	1910	ATTTTGAAGGCTTTTAAAC	1927
DB	621	IleLeuLysAlaPheAsn	626

RESULT 7
US-10-899-551-53
; Sequence 53, Application US/10899551
; Publication No. US20050063994A1
; GENERAL INFORMATION:
; APPLICANT: Caplan, Michael J.
; APPLICANT: Burks, A. Wesley
; APPLICANT: Sampson, Hugh A.
; APPLICANT: Howard, Sosin B.
; APPLICANT: Bottomly, Kim H.
; TITLE OF INVENTION: Methods and Reagents for Decreasing Clinical Reaction to Allergy
; FILE REFERENCE: 2002834-0233
; CURRENT APPLICATION NUMBER: US/10/899,551
; CURRENT FILING DATE: 2004-07-26
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.2

Db 297 PhePheProAlaSerSerArgAspGlnSerSerTyLeuGlnGlyPheSerArgAsnThr 316
Qy 1010 TTGGAGCGCCCTTCAATCGGAATTCATAGATACGAGGCTGTGTAGAGAGAAT 1069
Db 317 LeuGluAlaAlaPheAsnAlaGluPheAsnGluLeuArgValLeuLeuGluAsn 336
Qy 1070 GCAGGAGGTGAGCAAGAGGAGAGAGGCGAGCGATGAGTACTCGAGTAGTGAAC 1129
Db 337 AlaGlyGlyGluGlnGluArgGlyGlnArgTrpSerThrArgSerSerGluAsn 356
Qy 1130 AATGAAGGAGTGTAGTCAAGGTGCAAGGACGACGCTTCAAGCACTTACTAGACGCT 1189
Db 357 AsnGluGlyValLeuValLeuValSerLysGluHisValGluLeuThrLysHisAla 376
Qy 1190 AATCCGCTCTCAAGAAAGCTCCGAGAGAGGAGATATACCAACCCCAATCAACTTG 1249
Db 377 LysSerValSerLysLysGlySerGluGluGluGlyAspIleThrAsnProIleAsnLeu 396
Qy 1250 AGAAGGCGAGCCGATCTTTCTTAACAATTTGGGAAGTTATTGAGGTGAAGCCAGAC 1309
Db 397 ArgGluGlyGluProAspLeuSerAsnAsnPheGlyLysLeuPheGluValLysProAsp 416
Qy 1310 AAGAAGAACCCAGCTTCAGGACCTGACATGATCTACCTGTGTAGAGATCAAGAA 1369
Db 417 LysLysAsnProGlnLeuGlnAspLeuAspMetLeuThrCysValGluIleLysGlu 436
Qy 1370 GGAGCTTTGATGCTCCACACTTCAACTCAAGGCCATGTTATCGTCGTCAACAAA 1429
Db 437 GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValAsnLys 456
Qy 1430 GGAATCGAAACCTTGAACTCGTGGCTGAAGAAAGAGCAACAAGAGGGGACGGCG 1489
Db 457 GlyThrGlyAsnLeuGluLeuValAlaValLargLysGluGlnGlnArgGlyArg 476
Qy 1490 GAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1549
Db 477 GluGluGluGluAspGluAspGluGluGluGluGluGluGluGluGluGluGlu 496
Qy 1550 ACAGCGAGGTGAAGGAGCGGATGTGTTTCATCATGCGACAGCTCATCCAGTAGCCATC 1609
Db 497 ThrAlaArgLeuLysGluGlyAspValPheIleMetProAlaAlaHisProValAlaIle 516
Qy 1610 AACGTTCTCCGAACTCCATCTGTTGGCTTCGGTATCAACGCTGAAACCAACCAAGA 1669
Db 517 AsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyLeuAsnAlaGluAsnAsnHisArg 536
Qy 1670 ATCTTCTTCAGGTGATAGGACATGCTATAGACCGATAGACGAGGAGGAGGAGGAGGAG 1729
Db 537 IlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGlnAlaLysAsp 556
Qy 1730 TTAGCATTCCTCGGTGCGGTGAACAAAGTTGAGAAGCTCATCAAAACCAAGAGGAATCT 1789
Db 557 LeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGlnLysGluSer 576
Qy 1790 CACTTGTGAGTGTCTCTCAATCTCAATCTCAATCTCCGCTGCTCTCTGAGAAAGAG 1849
Db 577 HisPheValSerAlaArgProGlnSerGlnSerGlnSerProSerProGluLysGlu 596
Qy 1850 TCTCCTGAGAGGAGGATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1909
Db 597 SerProGluLysGluAspGlnGluGluGluAsnGlnGlyLysGlyProLeuLeuSer 616
Qy 1910 ATTTTGAAGGCTTTTAAAC 1927
Db 617 IleLeuLysAlaPheAsn 622

RESULT 8

US-09-731-221-78
; Sequence 78, Application US/09731221
; Patent No. US2002018778A1
; GENERAL INFORMATION:
; APPLICANT: Caplan, Michael

; TITLE OF INVENTION: Passive Desensitization
; FILE REFERENCE: 2002834-0103
; CURRENT APPLICATION NUMBER: US/09/731,221
; CURRENT FILING DATE: 2001-12-06
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 78
; LENGTH: 634
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Arachis
; OTHER INFORMATION: Hypogaea
US-09-731-221-78

Alignment Scores: 3,79e-242 Length: 634
Pred. No.: 3052.00 Matches: 584
Score: 96.86% Conservative: 3
Percent Similarity: 96.37% Mismatches: 19
Best Local Similarity: 85.68% Indels: 0
Query Match: 9 Gaps: 0
DB:

US-10-728-051-1 (1-2032) x US-09-731-221-78 (1-634)

Qy 110 TTGTCAACGCATGCCAAGTCATCACCCTTACCAAGAGAAACACAGAGAACCCCTCGGCCAG 169
Db 16 SerSerThrHisAlaLysSerSerProTyGlnAlaLysThrGluAsnProCysAlaGln 35
Qy 170 AGGTGCTCCAGAGTTGTCAACAGGAAACCGGATGACTTCAAGCANAAGCATCGAGTCT 229
Db 36 ArgCysLeuGlnSerCysGlnGluProAspAlaLeuLysGlnLysAlaCysGluSer 55
Qy 230 CGCTGCACCAAGCTCCAGTATGATCTCTGTGTCTATGATCTCTCGAGGACACACTGGC 289
Db 56 ArgCysThrLysLeuGluTyAspProArgCysAlaTyAspProArgGlyHisThrGly 75
Qy 290 ACCACCAACCAAGCTTCCCTCCAGGGAGCGGACACGTCGCCGCCCAACCCGGAGACTAC 349
Db 76 ThrThrAsnGlnArgSerProGlyGluAlaThrArgGlyArgGlnProGlyAspTyr 95
Qy 350 GATGATGACCGCTCAACCCCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 409
Db 96 AspAspAlaArgArgGlnProArgAlaGluGluGlyGlyArgTrpGlyProAlaGlyPro 115
Qy 410 AGGAGCGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 469
Db 116 ArgGluArgGluArgGluGluAspAlaArgGlnProArgGluAspTrpAlaArgProSer 135
Qy 470 CATCAGCAGCCCGGAAATAAGCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 529
Db 136 HisGlnGlnProArgLysAlaArgProGluGlyArgGluGlyGluGlnGluTrpGlyThr 155
Qy 530 CAGGTAGCCATCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 589
Db 156 ProGlySerHisValArgGluGluThrSerArgAsnAsnProPheTyPheProSerArg 175
Qy 590 CGGTTTAGCACCCGCTACCGGAAACCAACATCTCCGAGCAACCCCTTCTACTTCCCGTCAAGG 649
Db 176 ArgPheSerThrArgTyGlyAsnGlnAsnGlyArgIleArgValLeuGlnArgPheAsp 195
Qy 650 CAAAGTCAAGGAGTTTCAGATCTCCAGATCCACCGTATTCGTCAGATTCGAGGCCAAG 709
Db 196 GlnArgSerArgGlnPheGlnAsnLeuGlnAsnHisArgIleValGlnIleGluAlaLys 215
Qy 710 CCTAACACTCTTGTCTTCCAGAGCAGCTGATGCTGATGAATCATCTTGTATCCAGCAA 769
Db 216 ProAsnThrLeuValLeuProLysHisAlaAspAlaAspAsnIleLeuValIleGlnGln 235
Qy 770 GGGCAAGCCACCGTAGCCCTAGCAATGGCAATGGCAATGGCAATGGCAATGGCAATGGCAAT 829
Db 236 GlyGlnAlaThrValThrValAlaAsnGlyAsnAsnArgLysSerPheAsnLeuAspGlu 255

```
QY 830 GGCGATGCTCAGATCCCATCCGGTTTCATTCTCAGATCTTGAACCGCCATGACAC 889
Db 256 GlyHisAlaLeuArgIleProSerGlyPheIleSerTyrlleLeuAsnArgHisAspAsn 275
QY 890 CAGAACTCAGAGTAGCTAAATCTCCATGCGCGTTTAAACACCCCGCAGTTTGAGGAT 949
Db 276 GluAsnLeuArgValAlaIleSerMetProValAsnThrProGlyGlnMetGluAsp 295
QY 950 TTCTTCCCGCGAGCCGAGACCAATCATCTACTTCAGCGCTTCAGCGGATACG 1009
Db 296 PhePheProAlaSerArgAspGlnSerSerTyrlleGlnGlyPheAlaArgAsnThr 315
QY 1010 TTGGAGCCCGCTTCAATCGGAATCAATAGATACGAGGCTGCTTTAGAAAGAAAT 1069
Db 316 LeuGluAlaAlaPheAsnAlaGluAlaAsnGluIleArgArgValleuLeuGluAsn 335
QY 1070 GCAGGAGGTGAGCAAGAGGAGAGGCGAGCGATGAGTACTCGAGTGTAGTGAAC 1129
Db 336 AlaGlyGlyGluGlnGluAlaAArgGlyGlnArgArgTrpSerThrArgSerSerGluAsn 355
QY 1130 AATGAGGAGTGTAGTCAAGGTGTCAAAGGACGCTTGAAGAACTTACTAAGCACGCT 1189
Db 356 AsnGluGlyValIleValIleValIleValIleValIleValIleValIleValIle 375
QY 1190 AAATCCGCTCTCAAGAAAGCTCCGAGAGAGGAGATATCAACCAACCAATCAACTTG 1249
Db 376 LysSerValSerLysLysGlySerGluGluGluGlyAspIleThrAsnProAlaAsnLeu 395
QY 1250 AGAAGAGCGAGCGCGATCTTTCTAACACTTTGGGAAGTTATTGAGTGAAGCCAGAC 1309
Db 396 ArgGluGlyGluProAspLeuSerAsnAsnPheGlyLysLeuAlaGluValLysProAsp 415
QY 1310 AAGAAGAACCCAGCTTCAAGACCTGAGCATGATCTACCTGCTGTAGAGATCAAGAA 1369
Db 416 LysLysAsnProGlnLeuGlnAspLeuAspMetLeuThrCysValGluIleLysGlu 435
QY 1370 GGAGCTTTGATGCTCCACACTTCAACTCAAAGGCCATGTTATCGCTCGTCAACAA 1429
Db 436 GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValAsnLys 455
QY 1430 GGAACGGAACCTTGAACTCGTGGCTGTAAAGAAAGAGCAACACAGAGGGAGCGCG 1489
Db 456 GlyThrGlyAsnLeuGluLeuValAlaValArgLysGluGlnGlnArgGlyArgA-g 475
QY 1490 GAAGAAGAGAGAGCAAGACGAGAGAGAGAGAGAGTAAACAGAGAGTGCCTAGGTAC 1549
Db 476 GluGluGluGluAspGluAspGluGluGluGlySerAsnArgGluValArgAlaTy 495
QY 1550 ACAGCGAGGTGAAGGAGCGATGTGTTTCATCATGCCAGCAGCTCATCCAGTAGCCATC 1609
Db 496 ThrAlaArgLeuLysGluGlyAspValPheIleMetProAlaAlaHisProValAlaIle 515
QY 1610 AACGTTCTCCGNACTCCATCTGCTTGGCTTCGTTATCAACCTGAAACACACCCAGA 1669
Db 516 AsnAlaSerSerGluLeuAlaLeuLeuGlyPheGlyIleAsnAlaGluAsnAsnHisArg 535
QY 1670 ATCTTCTTCAGGTGATAAGGACAAATGTCATGACCCAGATAGAGAGCAAGCGAAGAT 1729
Db 536 IlePheLeuAlaGlyAspAlaAspAsnValIleAspGlnIleGluLysGlnAlaLysAla 555
QY 1730 TTAGCATTCCTGGTGGGTGAACAAGTTGAGAGCTCATCAAAACCAAGAGGATCT 1789
Db 556 LeuAlaAlaProGlySerGlyGluGlnValGluLysLeuIleLysAsnGlnLysGluSer 575
QY 1790 CACTTTCGAGTGTCTGCTCCTCAATCTCAATCTCCGCTGCTCCTCGAAGAAGAG 1849
Db 576 HisPheValAlaAlaAArgProGlnSerGlnSerProSerSerProGluLysGlu 595
QY 1850 TCTCCTGAGAAGAGGATCAAGAGGAGGAAACCAAGAGGGAGGGTCCACTCTTCA 1909
Db 596 SerProGluLysGluAspGlnGluGluAsnGlnGlyGlyLysGlyProLeuLeuSer 615
QY 1910 ATTTTGAAGGCTTTTAAAC 1927
```

```
Db 616 lleLeuLysAlaPheAsn 621
RESULT 9
US-09-331-631A-21
; Sequence 21, Application US/09331631A
; Patent No. US20020168392A1
; GENERAL INFORMATION:
; APPLICANT: Manners, John M.
; APPLICANT: Marcus, John Paul
; APPLICANT: Goulter, Kenneth C.
; APPLICANT: Green, Jodie L.
; FILE OF INVENTION: ANTIMICROBIAL PROTEINS
; FILE REFERENCE: CULLN23.001APC
; CURRENT APPLICATION NUMBER: US/09/331,631A
; CURRENT FILING DATE: 1999-06-21
; PRIOR APPLICATION NUMBER: PCT/AU97/00874
; PRIOR FILING DATE: 1997-12-22
; PRIOR APPLICATION NUMBER: AU PO 4275
; PRIOR FILING DATE: 1996-12-20
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: fastseq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 614
; TYPE: PRT
; ORGANISM: Peanut
US-09-331-631A-21
Alignment Scores:
Pred. No.: 3,01e-241 Length: 614
Score: 3041.00 Matches: 595
Percent Similarity: 96.03% Conservative: 9
Best Local Similarity: 94.59% Mismatches: 7
Query Match: 85.37% Indels: 18
DB: 9 Gaps: 7
US-10-728-051-1 (1-2032) x US-09-331-631A-21 (1-614)
QY 50 ATGAGAGGAGGAGGTTTCTCCACTGATGCTGTGTGTAGGATCCTTGTCTGTGCTTCAGTT 109
Db 1 MetArgGlyArgValSerProLeuMetLeuLeuGlyIleLeuValLeuAlaSerVal 20
QY 110 TCTGCAACGCATGCCAAGTCATCCTTACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 169
Db 21 SerAlaThrGlnAlaLys---SerProTyrl---ArgLysThrGluAsnProCysAlaGln 38
QY 170 AGTGCTCTCCAGATTGTCAACAGGAACCGGATGACTTGAAGCAAAAGCATGCGAGTCT 229
Db 39 ArgCysLeuGlnSerCysGlnGlnProAspAspLeuLysGlnLysAlaCysGluSer 58
QY 230 CGCTGCACCAAGCTCGAGTATGATCCTCGTTGTGTATGATCCTCGAGGACACACTGGC 289
Db 59 ArgCysThrLysLeuGluTyrlAspProArgCysValTyrlAsp-----ThrGly 74
QY 290 ACACCAACCAACGTTCCCTCCAGGGAGCGGACACGTCGGCGCCCAACCCGAGACTAC 349
Db 75 AlaThrAsnGlnArgHisProGlyGluArgThrArgGlyArgGlnProGlyAspTyrl 94
QY 350 GATGATGACCGCGCTCAACCCGGAAGAGAGAGGAGGCGGATGGGAGCCAGCTGGACCG 409
Db 95 AspAspAspArgArgGlnProArgArgGluGluGlyArgTrpGlyProAlaGluPro 114
QY 410 AGGAGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 469
Db 115 ArgGluArgGluArgGluAspTrpArgGlnProArgGluAspTrpArgArgProSer 134
QY 470 CATCAGCAGCCCGGAAATTAAGGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 529
Db 135 HisGlnGlnProArgLysIleArgProGluGlyArgGluGlyGluGluGlnIleTrpGlyThr 154
QY 530 CCAGGTAGCCATGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 589
Db 155 ProGlySerGluValArgGluThrSerArgAsnAsnProPheTyrlPheProSerArg 174
```

QY 590 CGTTTACGACCCGCTACGGGAACCAAAACGGTAGGATCCGGTCTCTGCAGAGGTTGAC 649
Db 175 ArgPheSerThrArgTyrGlyAsnGlnAsnGlyArgileArgValLeuGlnArgPheAsp 194
QY 650 CAAAGCTCAAGGAGTTCAGAAATCCAGAAATCACCGTATTGTGCAGATCGAGGCCAA 709
Db 195 GlnArgSerGlnPheGlnAsnLeuGlnAsnHisArgileValGlnIleGluAlaArg 214
QY 710 CCTAAACACTCTTCTTCCCAAGCACGCTGATGCTGATAACATCTTGTATTCCAGCAA 769
Db 215 ProAsnThrLeuValLeuProLysHisAlaAspAsnIleLeuValIleGlnGln 234
QY 770 GGCAAGCCACCTGACCGTAGCAATGCGCAATTAACAGAAAGCTTTAATCTTGCAG 829
Db 235 GlyGlnAlaThrValThrValAlaAsnGlyAsnAsnArgLysSerPheAsnLeuAspGlu 254
QY 830 GGCATGCACTCAGATCCCATCCGTTTCAATCTTCTTACATCTTGACCCGCTACAC 889
Db 255 GlyHisAlaLeuArgileProSerGlyPheIleSerTyrIleLeuAsnArgHisAspAsn 274
QY 890 CAGAACCTCAGAGTAGTAAATCTCCATCCCGCTTAAACACCCCGCCAGTTTGAGGAT 949
Db 275 GlnAsnLeuArgValAlaLysIleSerMetProValAsnThrProGlyGlnPheGluAsp 294
QY 950 TTCTTCCCGCGAGCCGAGACCAATCATCTCTTGCAGGGCTTCAGCAGGAATACG 1009
Db 295 PhePheProAlaSerSerArgAspGlnSerSerTyrLeuGlnGlyPheSerArgAsnThr 314
QY 1010 TTGAGGCCGCTTCAATCGGAAATCAATGAGATACGAGGCTGTGTTAGAAAGAT 1069
Db 315 LeuGluAlaAlaPheAsnAlaGluPheAsnGluIleArgArgValLeuLeuGluAsn 334
QY 1070 GCAGGAGTGCAGCAAGAGGAGGCGCAGAGCGATGGAGTACTCGGAGTAGGAGAAC 1129
Db 335 AlaGlyGlyGlnGlnGluArgGlyGlnArgSerThrArgSerSerAsp--- 353
QY 1130 AATCAAGGAGTGTAGTCAAGTGTCAAAGGACGCTTCAAGAACTTACTAAGACGCT 1189
Db 354 AsnGluGlyValIleValLysValSerLysGluHisValGlnGluLeuThrLysHisAla 373
QY 1190 AATCGTCTCAAGAAAGCTCGAAGAGAGGAGATATACCAACCCCAATCACTTG 1249
Db 374 LysSerValSerLysLysGlySerGluGluGlu---AspileThrAsnProIleAsnLeu 392
QY 1250 AGAAGAGCGGAGCCGATCTTCTCAACACTTTGGGAAGTATTTCAGGTGAGGCCAGAC 1309
Db 393 ArgAspGlyGluProAspLeuSerAsnAsnPheLysArgLeuPheGluValLysProAsp 412
QY 1310 AAGAAGAACCCCGCTTCAGGACCTGGACATGCTCACCTGTGTAGAGATCAAGAA 1369
Db 413 LysLysAsnProGlnLeuGlnAspLeuAspMetMetLeuThrCysValGluIleLysGlu 432
QY 1370 GAGCTTTGATGCTCCACACTTCACTCAAGGCCATGTTATCGTGTGCTCAACAA 1429
Db 433 GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValAsnLys 452
QY 1430 GGAACCTGGAACCTGGAACCTCGTGGCTGTAAGAAAGAGCAACACAGAGGGGCGCGG 1489
Db 453 GlyThrGlyAsnLeuGluLeuValAlaValArgLysGluGlnGlnArgGlyArg 472
QY 1490 GAA-----GAAGAGGAGGACGAAAGACGAGAGGAGGAGGAGTAAACAGAGAGGTG 1540
Db 473 GluGlnGlnLysGluGluGluGluAspGluGluGluGlySerAsnArgGluVal 492
QY 1541 CGTAGGTACACAGCGAGTTGAAGAGGCGATGTGTTTCATATGCGAGCGAGTCAATCCA 1600
Db 493 ArgArgTyrThrAlaArgLeuLysGluGlyAspValPheIleMetProAlaAlaHisPro 512
QY 1601 GTAGCCATCAACGCTTCTCCGAACTCCATCTGCTGGCTTCGTTATCAACGCTGAAC 1660
Db 513 ValAlaIleAsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsn 532

QY 1661 AACCAAGAAATCTTCTTCAGGTGATAGGACAATGTGATAGACAGATAGAGAACAA 1720
Db 533 AsnHisArgilePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGln 552
QY 1721 GCGAAGGATTAGCATTCCTCGGTGGTGAACAAGTTGAGAAGCTCATCAAAACCAG 1780
Db 553 AlaLysAspLeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGln 572
QY 1781 AAGGAATCTCACTTTGTGAGTGTCTCTCTCAATCTCAATCTCAATCTCTCGTCTCTCT 1840
Db 573 ArgGluSerHisPheValSerAlaArgProGlnSerGlnSerProSer----- 588
QY 1841 GAGAAAGAGTCTCTCAGAAAGAGGATCAAGAGGAGGAGAAACCAAGAGGAGGAGTCCA 1900
Db 589 -----SerProGlyLysGluAspGlnGluGluAsnGlnGlyGlyLysGlyPro 605
QY 1901 CTCCTTTCAATTTGAAGGCTTTTAAAC 1927
Db 606 LeuLeuSerIleLeuLysAlaPheAsn 614

RESULT 10
US-09-847-208-27
; Sequence 27, Application US/09847208
; Publication No. US20030082190A1
; GENERAL INFORMATION:
; APPLICANT: Saxon, Andrew
; APPLICANT: Zhang, Ke
; APPLICANT: Zhu, Daocheng
; TITLE OF INVENTION: FUSION MOLECULES AND TREATMENT OF
; TITLE OF INVENTION: IGH-MEDIATED ALLERGIC DISEASES
; FILE REFERENCE: UC67.002A
; CURRENT APPLICATION NUMBER: US/09/847,208
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 177
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 614
; TYPE: PRT
; ORGANISM: Arachis hypogaea (Peanut)
US-09-847-208-27

Alignment Scores:
Pred. No.: 3,01e-241 Length: 614
Score: 3041.00 Matches: 595
Percent Similarity: 96.03% Conservative: 9
Best Local Similarity: 94.59% Mismatches: 7
Query Match: 85.37% Indels: 18
DB: Gaps: 7

US-10-728-051-1 (1-2032) x US-09-847-208-27 (1-614)

QY 50 ATGAGAGGAGGAGGTTTCTCCACTGATGCTGTGCTAGGATCCTGTCTCGGTTCAAGTT 109
Db 1 MetArgLysArgValSerProLeuMetLeuLeuGlyIleLeuValLeuAlaSerVal 20
QY 110 TCTGCAACGCATGCCAAGTCATCACCTTACAGAGAAACACAGAACCCCTCGGCCAG 169
Db 21 SerAlaThrGlnAlaLys---SerProTyr---ArgLysThrGluAsnProCysAlaGln 38
QY 170 AGTGCTCTCAGAGTTGTCAACAGGAACCGGATGACTTGAGCAAAAGGATCGAGTCT 229
Db 39 ArgCysLeuGlnSerCysGlnGlnGluProAspAspLeuLysGlnLysAlaCysGluSer 58
QY 230 CGCTGCACCAAGCTCCAGTATGATCCTCGTTGTGTATGATCCTCGAGGACACACTGGC 289
Db 59 ArgCysThrLysLeuGluTyrAspProArgCysValTyrAsp-----ThrGly 74
QY 290 ACCACCAACCAACGTTTCCCTCCAGGGGAGCGGACACGTCGGCGGCCCAACCCCGAGACTAC 349
Db 75 AlaThrAsnGlnArgHisProProGlyGluArgThrArgGlyArgGlnProGlyAspTyr 94
QY 350 GATGATGACCGCGCTCAACCCCGAAGAGAGGAGGCGGATGGGGACCGCTGGACCG 409

Db 95 AspAspArgArgGlnProArgArgGluGluGlyGlyArgTrpGlyProAlaGluPro 114
QY 410 AGGAGCGTGAAGAGAGAAAGAACTCGAGACAACCAAGAGAAATGGAGGCGACCAAGT 469
Db 115 ArgGluArgGluArgGluAspTrpArgGlnProArgGluAspTrpArgProSer 134
QY 470 CATCAGAGCCAGCGAAATAAGGCCCGAAGAGAGAGAAAGAGAAACAAGAGTGGGAACA 529
Db 135 HisGlnGlnProArgLysIleArgProGluGlyArgGluGluGlnGlnTrpGlyThr 154
QY 530 CCAGGTAGCCATGTGAGGAGAGAAACATCTCGGAACAACCCCTTCTACTCCCGTCAAGG 589
Db 155 ProGlySerGluValArgGluGluThrSerArgAsnAsnProPheTyrPheProSerArg 174
QY 590 CGGTTTAGCACCCCGTACCGGAACCAACCGGTAGGATCCGGTCCCTGCAGAGGTTTGAC 649
Db 175 ArgPheSerThrArgTyrGlyAsnGlnAsnGlyArgIleArgValLeuGlnArgPheAsp 194
QY 650 CAAAGTCAAGGAGTTCAGAAATCTCCAGAAATCACCGTATTGTGAGATCCAGGCGCAAA 709
Db 195 GlnArgSerLysGlnPheGlnAsnLeuGlnAsnHisArgIleValGlnIleGluAlaArg 214
QY 710 CTTAAACACTCTGTTCTCCAGAGCAGCTGATGCTGATAACATCTTGTATCCAGCAA 769
Db 215 ProAsnThrLeuValLeuProLysHisAlaAspAlaAspAsnIleLeuValIleGlnGln 234
QY 770 GGGCAAGCCACCGTACCGGTAGCAAAATGGCAATTAACAGAAAGAGCTTTAATCTTGACGAG 829
Db 235 GlyGlnAlaThrValThrValAlaAsnGlnAsnArgLysSerPheAsnLeuAspGlu 254
QY 830 GGCCATGCACTCAGATCCCATCCGGTTTCATTTCTTACATCTTGAAACCGCCATGACAA 889
Db 255 GlyHisAlaLeuArgIleProSerGlyPheIleSerTyrIleLeuAsnArgHisAspAsn 274
QY 890 CAGAACTCAGAGTACGTAATACTCCATGCCCGTAAACACACCCCGCCAGTTTGAGGAT 949
Db 275 GlnAsnLeuArgValAlaLysIleSerMetProValAsnThrProGlyGlnPheGluAsp 294
QY 950 TTCTTCCCGCAGCAGCCGAGACCAATCATCTTCTGAGGCGTTCAGCAGGAATACG 1009
Db 295 PhePheProAlaSerSerArgAspGlnSerSerTyrLeuGlnGlyPheSerArgAsnThr 314
QY 1010 TTGGAGCGCCCTTCAATCGGAATTCATAGATACGAGGAGTGTCTGTAGAGAGAT 1069
Db 315 LeuGluAlaAlaPheAsnAlaGluPheAsnGluIleArgArgValLeuLeuGluGluAsn 334
QY 1070 GCAGGAGGTGAGCAGAGGAGCAGGCGAGCGATGGAGTACTCGGAGTAGTAGAAC 1129
Db 335 AlaGlyGlyGlnGluGluArgGlyGlnArgArgSerThrArgSerSerAsp--- 353
QY 1130 AATGAAGGAGTGTAGTCAAAAGTGTCAAAGGACGAGCTTGAAGAACTTACTAAGCAGCT 1189
Db 354 AsnGluGlyValIleValLysValSerLysGluHisValGlnGluLeuThrLysHisAla 373
QY 1190 AATCCGTCTCAAAGAAAGCTCCGAGAAAGAGGAGATATCAACCAACCAATCAACTTG 1249
Db 374 LysSerValSerLysLysGlySerGluGluGlu---AspIleThrAsnProIleAsnLeu 392
QY 1250 AGAAGAGGCGAGCCGATCTTCTTAACACTTTGGGAAGTATTTCAGGTGAGCGAGAC 1309
Db 393 ArgAspGlyGluProAspLeuSerAsnAsnPheGlyArgLeuPheGluValLysProAsp 412
QY 1310 AAGAAGAACCCCGAGCTTACGAGACCTGGACATGATGCTACCTGTGTAGAGATCAAGAA 1369
Db 413 LysLysAsnProGlnLeuGlnAspLeuAspMetMetLeuThrCysValGluIleLysGlu 432
QY 1370 GGAGCTTTGATGCTCCACACTTCAACTCAAGGCGCATGTTATCGTGTGCTCAACAA 1429
Db 433 GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValValAsnLys 452
QY 1430 GGAACGTGAACCTTGAACCTCGTGGCTGTGAAGAAAGAGCAACAACAGAGGCGCGCG 1489
Db 453 GlyThrGlyAsnLeuGluLeuValAlaValArgLysGlnGlnGlnArgGlyArgArg 472

QY 1490 GAA-----GAAGAGGAGGACGACGAGACGAGAGGAGGAGTAAACAGAGAGTGG 1540
Db 473 GluGlnGluTrpGluGluGluGluAspGluGluGluGluGlySerAsnArgGluVal 492
QY 1541 CGTAGGTACACAGCAGGTTGAAGAGGCGATGTTTCATCATGCCAGCAGCTCATCCA 1600
Db 493 ArgArgTyrThrAlaArgLeuLysGluGlyAspValPheIleMetProAlaAlaHisPro 512
QY 1601 GTAGCCATCAACGCTTCCCGAACTCCATCTGCTTGGCTTCGGTATCAACGCTGAAAC 1660
Db 513 ValAlaIleAsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsn 532
QY 1661 AACCCAGCAATCTCTTCAGGTGATAAGGACAATGTGATAGACAGATAGAGAGCAA 1720
Db 533 AsnHisArgIlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGln 552
QY 1721 GCGAGGATTTAGCATTCCTCGGTCCGGTGAAGTGAAGTGAAGTCAATAAACACAG 1780
Db 553 AlaLysAspLeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGln 572
QY 1781 AAGGAATCTCACTTTGTGAGTCTCTCTCAATCTCAATCTCAATCTCCGTCGTCCT 1840
Db 573 ArgGluSerHisPheValSerAlaArgProGlnSerGlnSerProSer----- 588
QY 1841 GAGAAAGATCTCTGAGAAAGAGGATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGTCCA 1900
Db 589 -----SerProGluLysGluAspGlnGluGluAsnGlnGlyGlyLysGlyPro 605
QY 1901 CTCCTTCAATTTGAAGCTTTTAAC 1927
Db 606 LeuLeuSerIleLeuLysAlaPheAsn 614

RESULT 11
US-10-147-095-21
; Sequence 21, Application US/10147095
; Publication No. US20030171274A1
; GENERAL INFORMATION:
; APPLICANT: Manners, John M.
; APPLICANT: Marcus, John Paul
; APPLICANT: Goulter, Kenneth C.
; APPLICANT: Green, Jodie L.
; TITLE OF INVENTION: ANTIMICROBIAL PROTEINS
; FILE REFERENCE: CULN23.001APC
; CURRENT APPLICATION NUMBER: US/10/147,095
; CURRENT FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US/09/331,631A
; PRIOR FILING DATE: 1999-06-21
; PRIOR APPLICATION NUMBER: PCT/AU97/00874
; PRIOR FILING DATE: 1997-12-22
; PRIOR APPLICATION NUMBER: AU PO 4275
; PRIOR FILING DATE: 1996-12-20
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 614
; TYPE: PRT
; ORGANISM: Peanut
US-10-147-095-21

Alignment Scores:
Pred. No.: 3,01e-241 Length: 614
Score: 3041.00 Matches: 595
Percent Similarity: 96.03% Conservative: 9
Best Local Similarity: 94.59% Mismatches: 7
Query Match: 85.37% Indels: 18
DB: 14 Gaps: 7

US-10-728-051-1 (1-2032) x US-10-147-095-21 (1-614)

QY 50 ATGAGAGGAGGAGGTTTCTCCACTGATGCTGTGTAGGGATCCTTGTCTCGTTCAGTT 109
Db 1 MetArgGlyArgValSerProLeuLeuLeuGlyIleLeuValLeuAlaSerVal 20

```
QY 110 TCTGCAACCATCCAGTCACTACCTTACCAGAAAGAAACAGAACCCCTCGCCGAC 169
Db 21 SerAlaThrGlnAlaLys---SerProTyr---ArgLysThrGluAsnProCysAlaGln 38
QY 170 AGTGCTCCAGAGTGTCAACAGGAACCGGATGACTTGAAGCAAAAGGCGATCCGAGTCT 229
Db 39 ArgCysLeuGlnSerCysGlnGlnGluProAspAspLeuLysGlnLysAlaCysGluSer 58
QY 230 CGTGCAACCAAGTCCAGTATGATCTCGTGTGTCTATGATCCTCGAGGACACACTGGC 289
Db 59 ArgCysThrLysLeuGlnTyrAspProArgCysValTyrAsp-----ThrGly 74
QY 290 ACCACCAACCAAGCTTCCCTCCAGGGAGCGGACGCTGCGCCCAACCCGAGACTAC 349
Db 75 AlaThrAsnGlnArgHisProProGlyGluArgThrArgGlyArgGlnProGlyAspTyr 94
QY 350 GATGATGACCCGCTCAACCCCAAGAGAGAGGAGGCGGATGGGGACCCAGCTGGACCG 409
Db 95 AspAspAspArgArgGlnProArgArgGluGluGlyGlyArgTyrProAlaGluPro 114
QY 410 AGGAGCGTGAAGAAGAACTCGAGACAAACCAAGAGAAAGATTGGAGGCGACCAAGT 469
Db 115 ArgGluArgGluArgGluAspTyrArgGlnProArgGluAspTyrArgProSer 134
QY 470 CATCAGCCAGCCGAAATAAGCCCGAAGGAAGAGAGAGGAAACAAGATGGGGAAACA 529
Db 135 HisGlnGlnProArgLysIleArgProGluGlyArgGluGlyGluGlnTyrGlyThr 154
QY 530 CAGGTTAGCATGTGAGGAAGAAACATCTCGGAACAACCCCTTCTACTCCGCTCAAGG 589
Db 155 ProGlySerGluValArgGluGluThrSerArgAsnAsnProPheTyrPheProSerArg 174
QY 590 CGGTTTAGCACCCGCTACCGGAACCAAAACCGTAGGATCCGGTCTCGAGAGGTTTGAC 649
Db 175 ArgPheSerThrArgTyrGlyAsnGlnAsnGlyArgIleArgValLeuGlnArgPheAsp 194
QY 650 CAAAGTCAAGGAGTTCAGAAATCCAGAAATCCAGAAATCCAGTATTGTGAGTCCAGGCCAAA 709
Db 195 GlnArgSerLysGlnPheGlnAsnLeuGlnAsnHisArgIleValGlnIleGluAlaArg 214
QY 710 CTTAAACACTCTGTTCTCCAGACGCTGATGCTGATTAACATCTGTTATCCAGCAA 769
Db 215 ProAsnThrLeuValLeuProLysHisAlaAspAlaAspAsnIleLeuValIleGlnGln 234
QY 770 GGGCAAGCCACCGTACCGTAGCAAAATGGCAATAACAGAAAGAGCTTTAATCTTGACGAG 829
Db 235 GlyGlnAlaThrValThrValAlaAsnGlyAsnAsnArgLysSerPheAsnLeuAspGlu 254
QY 830 GGCCATGCACTCAGATCCCATCCGTTTCATTTCTTACATCTTGAACCGCCCATGACAAC 889
Db 255 GlyHisAlaLeuArgIleProSerGlyPheIleSerTyrIleLeuAsnArgHisAspAsn 274
QY 890 CAGAACTCAGATAGTAAATCTCCATCCCGCTTAAACACACCCCGCCGCTTTAGGAT 949
Db 275 GlnAsnLeuArgValAlaLysIleSerMetProValAsnThrProGlyGlnPheGluAsp 294
QY 950 TTCTTCCCGCAGCAGCGAGACCAATCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1009
Db 295 PhePheProAlaSerArgAspGlnSerSerTyrLeuGlnGlyPheSerArgAsnThr 314
QY 1010 TTGGAGGCGCCCTCAATCGGAATTCATAGATACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1069
Db 315 LeuGluAlaAlaPheAsnAlaGluPheAsnGluIleArgArgValLeuLeuGluGluAsn 334
QY 1070 GCAGGAGGTGAGCAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1129
Db 335 AlaGlyGlyGluGlnGluArgGlyGlnArgArgSerThrArgSerSerAsp--- 353
QY 1130 AATGAGGAGGTAGTCAAGTGTCAAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1189
Db 354 AsnGluGlyValIleValLysValSerLysGluHisValGlnGluLeuThrLysHisAla 373
```

```
QY 1190 AAATCGTCTCAAGAAAGCTCCGAAAGAGGAGGAGATATCACCACCAATCAACTTG 1249
Db 374 LysSerValSerLysLysGlySerGluGluGlu---AspIleThrAsnProIleAsnLeu 392
QY 1250 AGAGAAGCGGAGCCCGATCTTTCAACACTTTGGGAAGTATTTCAGGTGAAGCCAGAC 1309
Db 393 ArgAspGlyGluProAspLeuSerAsnAsnPheGlyArgLeuPheGluValLysProAsp 412
QY 1310 AGAGAAGACCCCGAGCTTCAGGACCTGGACATGATCTCCTCTGTGTAGAGATCAAGAA 1369
Db 413 LysLysAsnProGlnLeuGlnAspLeuAspMetMetLeuThrCysValGluIleLysGlu 432
QY 1370 GGAGCTTTCATCTCCACACTTCAACTCAAGGCCATGGTTATCTCTCGTCAACAAA 1429
Db 433 GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValValAsnLys 452
QY 1430 GGAATCGGAAACTTGAACCTCGTGGCTGTAAGAAAGAGCAACACAGAGGGGACGCGG 1489
Db 453 GlyThrGlyAsnLeuGluValAlaValArgLysGluGlnGlnGlnArgGlyArgArg 472
QY 1490 GAA-----GAAGAGGAGGACGAAAGACGAAAGAGAGGAGGAGGAGGAGGAGG 1540
Db 473 GluGlnGlnTyrGluGluGluGluGluGluGluGluGluGluGluGluGluVal 492
QY 1541 CGTAGGTACACAGCGAGGTTGAAGGAGGCGATGTGTTTCATCATGCCAGCAGCTCATCCA 1600
Db 493 ArgArgTyrThrAlaArgLeuLysGluGlyAspValPheIleMetProAlaAlaHisPro 512
QY 1601 GTAGCCATCAACGCTTCCCGAACTCCATCTCGTGGTTCGTTATCAACGCTGAAAC 1660
Db 513 ValAlaIleAsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsn 532
QY 1661 AACCCAGAGATCTTCTTCAGGTGATAAGGACAATGTGATAGACAGATAGAGAGCAA 1720
Db 533 AsnHisArgIlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGln 552
QY 1721 GCGAAGGATTTAGCATTCCTCGGTGCGGTGAACAAGTTGAGAGCTCATCAAAAACAG 1780
Db 553 AlaLysAspLeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGln 572
QY 1781 AAGGAATCTCCTTGTGAGTGTCTGCTCAATCTCAATCTCAATCTCTCGTCTCTCT 1840
Db 573 ArgGluSerHisPheValSerAlaArgProGlnSerGlnSerProSer----- 588
QY 1841 GAGAAAGAGTCTCTGAGAAAGAGGATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1900
Db 589 -----SerProGluLysGluAspGlnGluGluGluGluGluGluGlyGlyLysGlyPro 605
QY 1901 CTCCTTTCAATTTTGAAGCTTTTAAAC 1927
Db 606 LeuLeuSerIleLeuLysAlaPheAsn 614
RESULT 12
US-10-100-303A-8
; Sequence 8, Application US/10100303A
; Publication No. US20030202980A1
; GENERAL INFORMATION:
; APPLICANT: Caplan, et al.
; TITLE OF INVENTION: Methods and Reagents for Decreasing Clinical Reaction
; TITLE OF INVENTION: to Allergy
; FILE REFERENCE: 2002834-0166
; CURRENT APPLICATION NUMBER: US/10/100,303A
; CURRENT FILING DATE: 2002-03-18
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 614
; TYPE: PRT
; ORGANISM: Atachis hypogaea, Prot/Nucleo-Ara h 1
US-10-100-303A-8
Alignment Scores: 3.01e-241 Length: 614
Pred. No.:
```


Score:	3041.00	Matches:	595
Percent Similarity:	96.03%	Conservative:	9
Best Local Similarity:	94.59%	Mismatches:	7
Query Match:	85.37%	Indels:	18
DB:	15	Gaps:	7
US-10-728-051-1 (1-2032) x US-10-100-303A-8 (1-614)			
QY	50	ATGAGAGGAGGGTTTCTCCACTGATGCTGTCTAGGATCCTTGTCTCGCTTCAGTT	109
Db	1	MetArgGlyArgValSerProLeuMetLeuLeuLeuGlyIleLeuValLeuAlaSerVal	20
QY	110	TTCTGCACGATCCCAAGTCATCACCTTACAGAAGAAACAGAAACCCCTCGCCAG	169
Db	21	SerAlaThrGlnAlaLys---SerProTyr---ArgLysThrGluAsnProCysAlaGln	38
QY	170	AGTGCTCCAGAGTTGTCAACAGGAACCGGATGACTTGAAGCAAAAGGATCGAGTCT	229
Db	39	ArgCysLeuGlnSerCysGlnGlnGluProAspAspLeuLysGlnLysAlaCysGluSer	58
QY	230	CGCTGCACCAAGCTTCGAGTATGATCCTCGTTGTCTATGATCCTCGAGGACACACTGGC	289
Db	59	ArgCysThrLysLeuGluTyrAspProArgCysValTyrAsp-----ThrGly	74
QY	290	ACCACCAACCAAGTTCCTCCAGGGAGCGGACACGTCGGCCGCCAACCCGGAGACTAC	349
Db	75	AlaThrAsnGlnArgHisProProGlyGluArgThrArgGlyArgGlnProGlyAspTyr	94
QY	350	GATGATGACCGCGCTCAACCCCGAGAGAGAGGAGGCGGATGGGACCAAGCTGGACCG	409
Db	95	AspAspAspArgArgGlnProArgArgGluGluGlyArgTrpGlyProAlaGluPro	114
QY	410	AGGAGCGTCAAAAGAGAAGAAGCTCGAGACCAACCAAGAGAAATGGAGGCGCAACAGT	469
Db	115	ArgGluArgGluArgGluAspTrpArgGlnProArgGluAspTrpArgProSer	134
QY	470	CATCAGCAGCCAGAAAATAAGCCCGAAGAGAGAGAGAGAAACAAAGATGGGGAACA	529
Db	135	HisGlnGlnProArgLysIleArgProGluGlyArgGluGlyGluGlnGluTrpGlyThr	154
QY	530	CCAGGTAGCATGTGAGGAGAAACATCTCGAAACACCCCTTCTACTTCCGCTCAAG	589
Db	155	ProGlySerGluValArgGluGluThrSerArgAsnAsnProPheTyrPheProSerArg	174
QY	590	CGGTTTAGCACCCGCTACGGGAACCAAAACCGTAGGATCCGGTCTCGAGAGTTTGAC	649
Db	175	ArgPheSerThrArgTyrGlyAsnGlnAsnGlyArgIleArgValLeuGlnArgPheAsp	194
QY	650	CAAGGTCAAGCGATTTCAGAAATCTCCAGATCACCCTATTGTGAGATCGAGGCCAAA	709
Db	195	GlnArgSerLysGlnPheGlnAsnLeuGlnAsnHisArgIleValGlnIleGluAlaArg	214
QY	710	CCTAACACTTGTCTTCCAGCAGCAGCTGATGCTGATAACATCCTTGTATTCCAGCAA	769
Db	215	ProAsnThrLeuValLeuProLysHisAlaAspAlaAspAsnIleLeuValIleGlnGln	234
QY	770	GGCAAGCCACCGTAGCCGTAGCAAAATGCAATAACAGAAAGAGCTTTAATCTTGACGAG	829
Db	235	GlyGlnAlaThrValThrValAlaAsnGlyAsnAsnArgLysSerPheAsnLeuAspGlu	254
QY	830	GGCCATGCACTAGAAATCCCATCCCGTTTCATTTCTTACATCTTGAACCCGCCATGACAC	889
Db	255	GlyHisAlaLeuArgIleProSerGlyPheIleSerTyrIleLeuAsnArgHisAspAsn	274
QY	890	CAGAACTCAGAGTAGTAAATCTCCATGCCCGTTTAAACACACCCCGCCAGTTTGGAT	949
Db	275	GlnAsnLeuArgValAlaLysIleSerMetProValAsnThrProGlyGlnPheGluAsp	294
QY	950	TTCTTCCCGGAGCAGCCGAGACCAATCATCTTCTTGCAGGCGCTTCAGCAGGAATACG	1009
Db	295	PhePheProAlaSerArgAspGlnSerSerTyrLeuGlnGlyPheSerArgAsnThr	314
QY	1010	TTGGAGCCCGCTTCAATCGGAAATCAATGAGATACGAGGGTGCTGTAGAGAGAAT	1069

Db	315	LeuGluAlaAlaPheAsnAlaGluPheAsnGluIleArgValLeuLeuGluGluAsn	334
QY	1070	GCAGGAGGTGAGCAAGAGGAGAGGGCAGAGGGATGGAGTACTCTCGAGTGTAGTGAAC	1129
Db	335	AlaGlyGlyGluGlnGluArgGlyGlnArgArgSerThrArgSerAsp---	353
QY	1130	AATGAAGGAGTGATAGTCAAAAGTGTCAAAGGAGACGCTTGAAGAACTTACTTAAGACGCT	1189
Db	354	AsnGluGlyValIleValLysValSerLysGluHisValGlnGlnLeuThrLysHisAla	373
QY	1190	AAATCGCTTCAAAGAAAGGCTCCGAGAAGAGGAGATATCAACCAACCAATCAACTTG	1249
Db	374	LysSerValSerLysLysGlySerGluGluGlu---AspIleThrAsnProIleAsnLeu	392
QY	1250	AGAGAAGCGAGCCCGATCTTCTAAACAACCTTGGGAAGTTATTGTAGGTGACAGCCAGAC	1309
Db	393	ArgAspGlyGluProAspLeuSerAsnAsnPheGlyArgLeuPheGluValLysProAsp	412
QY	1310	AAGAAAGAACCCCGACTTCAGACCTCGACATGATGTCTCACTGTGTAGAGATCAAAAGAA	1369
Db	413	LysLysAsnProGlnLeuGlnAspLeuAspMetMetLeuThrCysValGluIleLysGlu	432
QY	1370	GGAGCTTTGATGCTCCCACTTCAACTCAAAGGCCATGGTTATTCGTCTGCTCAACAA	1429
Db	433	GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValAsnLys	452
QY	1430	CGAATCTGGAACCTTGAATCTGTGCTGTAGAAAGACCAACACAGAGGGGACGCGCG	1489
Db	453	GlyThrGlyAsnLeuGluValAlaValArgLysGluGlnGlnArgGlyArgArg	472
QY	1490	GAA-----GAAGAGGAGGACGACCAAGACGAGGAGGAGGAACTAAACAGAGAGTG	1540
Db	473	GluGlnGluTrpGluGluGluAspGluGluGluGlySerAsnArgGluVal	492
QY	1541	CTAGGTCACACGAGGTTGAAGAAAGCGATGTTCATCATGCGCAGCAGCTCATCCA	1600
Db	493	ArgArgTyrThrAlaArgLeuLysGluGlyAspValPheIleMetProAlaAlaHisPro	512
QY	1601	GTAGCCATCAACGCTTCTCCGAACTCCATCTGCTTGGCTTCGTTATCAACGCTGAAAC	1660
Db	513	ValAlaIleAsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsn	532
QY	1661	AACCAACAATCTTCTTGCAGGTGATAAGGACAAATGTATAGACCAAGATAGAGAAGCAA	1720
Db	533	AsnHisArgIlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGln	552
QY	1721	GGCAAGGATTAGCATTCCTCGGGTCCGGTGGAACAGTTGAGAAGCTCATCAAAACCAG	1780
Db	553	AlaLysAspLeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGln	572
QY	1781	AAGGAATCTCATTTGTAGTGCCTCGCTCAATCTCAATCTCAATCTCCGTCGCTCCT	1840
Db	573	ArgGluSerHisPheValSerAlaArgProGlnSerGlnSerProSer-----	588
QY	1841	GAGAAAGATCTCTCCGAAAGAGGATCAAGACGAGGAGAAACCAAGAGGAGGAGGGTCCA	1900
Db	589	-----SerProGluLysGluAspGlnGluAsnGlnGlyLysGlyPro	605
QY	1901	CTCCTTTCAATTTTCAAGGCTTTTAAAC	1927
Db	606	LeuLeuSerIleLeuLysAlaPheAsn	614

RESULT 13

US-10-100-303A-55

Sequence 55, Application US/10100303A

Publication NO. US20030202980A1

GENERAL INFORMATION:

APPLICANT: Caplan, et al.

TITLE OF INVENTION: Methods and Reagents for Decreasing Clinical Reaction

TITLE OF INVENTION: to Allergy

FILE REFERENCE: 2002834-0166

CURRENT APPLICATION NUMBER: US/10/100.303A

RESULT 13
US-10-100-303A-55
; Sequence 55, Application US/10100303A
; Publication No. US20030202980A1
; GENERAL INFORMATION:
; APPLICANT: Caplan, et al.
; TITLE OF INVENTION: Methods and Reagents for Decreasing Clinical Reaction
; TITLE OF INVENTION: to Allergy
; FILE REFERENCE: 2002834-0166
; CURRENT APPLICATION NUMBER: US/10/100,303A

QY	50	ATGAGAGGAGGGTTTCTCCACGTAGTCTGTCTAGGATCCTTGTCTCGCTTCAGTT	109
DB	2	MetArgAlaArgPhePro-----LeuLeuLeuLeuGlyLeuValPheLeuAlaSerVal	19
QY	110	TCTGCAACGCATCCCAAGTCATCACCATTACAGAGAAGAAACAGAGAACCCCTCGGCCACG	169
DB	20	SerValSerPheGlyIleAla-----TyrTrpGluLys--GluAsnProLysHisAsn	36
QY	170	AGGTGCTCTCCAGAGTTGTCAACAGGAACCGGAGTACTTGAAGCAAAAGCATCGCGAGTCT	229
DB	37	LysCysAsnLeuGlnSerCysAsnSerGluArgAspSerTyrArgAsnGlnAlaCysHisAla	56
QY	230	CGCTGCACCAAGCTCGAGTATGAT-----CCTCGTTGT	262
DB	57	ArgCysAsnLeuLeuLysValGluLysGluGluCysGluGluGlyGluIleProArg---	75
QY	263	GTCATTATGATCTCGAGGACACACTGGCCACCAACCAACAGTTCCTCCAGGGGACGG	322
DB	76	-----ProArgProArg-----ProGlnHisProGlu	84
QY	323	ACAGTGGCCGCCAACCCGGAGACTACGATGATGAC-----	358
DB	85	ArgGluProGlnGlnProGlyGluLysGluGluAspGluAspGluGlnProArgProfile	104
QY	359	-----CGCCGTCAACCCCGAAGAG-----GAAGGAGGCCGA	391
DB	105	ProPheProArgProGlnProArgGlnGluGluHisGluGlnArgGluGlnGlu	124
QY	392	TGG-----GGACCAAGCTGGACCGAGGCGGTCAAAGAGAAGAA	430
DB	125	TrpProArgLysGluGluLysArgGlyGluLysGlySerGluGluGluAspGluAspGlu	144
QY	431	GACTGGAGACACCAAGAGAAGAT-----TGGAGGCGACCAAGTCATCAGCAGCCACGG	484
DB	145	AspGluGluGlnAspGluArgGlnPheProPheProArgProProHisGlnLysGluGlu	164
QY	485	AAAAATAAGCCCGAAGGAGAGAGGAGAACAAAGATGGGGAAACACCGGTAGCCCATGTG	544

Db 165 ArgGlyGlnGluAspGluAspGluGlnGlnArgGluSerGluGluSerGluAsp 184
QY 545 AGGAA-----GAACATCTCGAACAACCCCTTTCTACTTCCCGTCAAGCGGTTTAGC 598
Db 185 SerGluLeuArgArgHisLysAsnLysAsnProPheLeuPheGlySerAsnArgPheGlu 204
QY 599 ACCGGTACGGGAACCAACCGTAGATCCGGGTCTCGAGAGTTTGACCAAGGTCA 658
Db 205 ThrLeuPheLysAsnGlnTyrGlyArgIleArgValLeuGlnArgPheAsnGlnArgSer 224
QY 659 AGCAGTTTCAGATCTCCAGATACCCGTATTGTGCAGATCAGGCCAAACCTACACT 718
Db 225 ProGlnLeuGlnAsnLeuArgAspTyrArgIleGluPheAsnSerLysProAsnThr 244
QY 719 CTGTCTTCCCAAGCAGCGTGTATGCTGATAACATCTGTATTATCCAGCAAGGCAAGCC 778
Db 245 LeuLeuLeuProAsnHisAlaAspAlaAspTyrLeuIleValIleLeuAsnGlyThrAla 264
QY 779 ACCGTGACCGTAGCAATGCAATACAGAAAGAGCTTTAATCTTACGAGGCCCATGCA 838
Db 265 IleLeuSerLeuValAsnAsnAspArgAspSerTyrArgLeuGlnSerGlyAspAla 284
QY 839 CTCAGATCCCATCCGTTTCATTTCTTACATCTTGAACCGCCATCACACAGAACCTC 898
Db 285 LeuArgValProSerGlyThrThrTyrTyrValValAsnProAspAsnGluAsnLeu 304
QY 899 AGAGTAGTAAATCTCCATGCCCGTTAAACACACCGCGGCGAGTTTGAGGATTTCTCCCG 958
Db 305 ArgLeuIleThrLeuAlaIleProValAsnLysProGlyArgPheGluSerPheLeu 324
QY 959 GCAGCAGCGAGCAACCATCTCTTGCAGGGTTTCAGCAGGAATACGTTGGAGGCC 1018
Db 325 SerSerThrGluAlaGlnGlnSerTyrLeuGlnGlyPheSerArgAsnIleLeuGluAla 344
QY 1019 GCCTTCAATCGGAATCAATGAGATACCGAGGCTCTGTAGAGAGATGACAGAGGT 1078
Db 345 SerTyrAspThrLysPheGluGluIleAsnLysValLeuPheSerArgGluGluGly--- 363
QY 1079 GAGCAAGAGAGAGAGCGGAGCGGATGAGTACTCGGAGTAGTGAGAACATGAAGGA 1138
Db 364 -----GlnGlnGlnGlyGlnArgLeu-----GlnGluSer 374
QY 1139 GTGATAGTCAAGTGTCAAGAGAGCGCTTGAAGAACTTACTAAGCAGCGTAAATCCGTC 1198
Db 375 ValIleValGluIleSerLysGluGlnIleArgAlaLeuSerLysArgAlaLysSerSer 394
QY 1199 TCAAGAAAGGCTCCGAAGAGAGGAGATATCACCAACCATCACTTGTAGAGAGGC 1258
Db 395 SerArgLysThrLysSerSerGlu-----AspLysProPheAsnLeuArgSerArg 411
QY 1259 GAGCCGATCTTTCTAACAACTTTGGGAAGTTATTGAGGTGAAGCCAGACAGAGAAGAAC 1318
Db 412 AspProIleTyrSerAsnLysLeuGlyLysPhePheGluIleThrProGlu---LysAsn 430
QY 1319 CCCAGCTTCAGACCTGACATGATGCTCCTCCTGTGTAGATCAAGAGAGCGTTTG 1378
Db 431 ProGlnLeuArgAspLeuAspIlePheLeuSerIleValAspMetAsnGluGlyAlaLeu 450
QY 1379 ATGCTCCCACTTCAACTCAAGGCGCATGTTATCGTCTGCTCAACAAAGGAAGTGA 1438
Db 451 LeuLeuProHisPheAsnSerLysAlaIleValIleLeuValIleAsnGluGlyAspAla 470
QY 1439 AACCTTTGAATCTGTGCTGTGAAGAAAGAGCAACAGAGGGGAGCGGGGAAGAGAG 1498
Db 471 AsnIleGluLeuValGlyLeu---LysGluGlnGlnGlnGlnGlnGlnGln 489
QY 1499 GAGGACGAGACCAAGAGAGGAGGAAGTAACAGAGAGTGCGTAGGTACACAGGAGG 1558
Db 490 Pro-----LeuGluValArgLysTyrArgAlaGlu 499
QY 1559 TTGAAGGAGCGGATGTTTCATCGCCAGCAGCTCATCCAGTAGCTCAACGCTTCC 1618
Db 500 LeuSerGluGlnAspIlePheValIleProAlaGlyTyrProValValValAsnAlaThr 519

QY 1619 TCCGAATCCATCTGCTTGGCTTCGATCAACGCTGAAAAACACACAGAAATCTTCCCTT 1678
Db 520 SerAsnLeuAsnPhePheAlaIleGlyIleAsnAlaGluAsnAsnGlnArgAsnPheLeu 539
QY 1679 GAGGTGATAAGCAATGTGTAGACCAAGATAGAGAACGAGCGAGGATTTAGCATTC 1738
Db 540 AlaGlySerGlnAspAsnValIleSerGlnIleProSerGlnValGlnGluLeuAlaPhe 559
QY 1739 CTGGGTCCGGTGAACAGTTGAGAGCTCATCAAAACCAAGCAAGAAATCTCACTTTGTG 1798
Db 560 ProGlySerAlaGlnAlaValGluLysLeuLysAsnGlnArgGluSerTyrPheVal 579
QY 1799 AGTGCTCGCTCAATCTCAATCTCGCTGCTCTCTCGAGAAAGAGTCTCTCTGAG 1858
Db 580 AspAlaGlnProLysLys----- 585
QY 1859 AAAGAGGATCAAGAGAGAGAAACCAAGAGGAGGAGGTCCTCTTCAATTTTGAAG 1918
Db 586 -----LysGluGluGlyAsnLysGlyArgLysGlyProLeuSerSerIleLeuArg 602
QY 1919 GCTTTT 1924
Db 603 AlaPhe 604
RESULT 15
US-09-331-631A-25
; Sequence 25, Application US/09331631A
; Patent No. US20020168392A1
; GENERAL INFORMATION:
; APPLICANT: Manners, John M.
; APPLICANT: Marcus, John Paul
; APPLICANT: Goulter, Kenneth C.
; APPLICANT: Green, Jodie L.
; TITLE OF INVENTION: ANTIMICROBIAL PROTEINS
; FILE REFERENCE: CULN23.001AEC
; CURRENT APPLICATION NUMBER: US/09/331,631A
; PRIOR FILING DATE: 1999-06-21
; PRIOR APPLICATION NUMBER: PCT/AU97/00874
; PRIOR FILING DATE: 1997-12-22
; PRIOR APPLICATION NUMBER: AU PO 4275
; PRIOR FILING DATE: 1996-12-20
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 25
; LENGTH: 605
; TYPE: PRT
; ORGANISM: Soybean (Glycine max)
US-09-331-631A-25
Alignment Scores:
Pred. No.: 2,49e-96 Length: 605
Score: 1281.00 Matches: 285
Percent Similarity: 59.21% Conservative: 107
Best Local Similarity: 43.05% Mismatches: 174
Query Match: 35.96% Indels: 96
DB: 9 Gaps: 18
US-10-728-051-1 (1-2032) x US-09-331-631A-25 (1-605)
QY 50 ATGAGAGGAGGAGGTTTCTCCATGATGCTGTGTAGGATCCTTGTCTCGGCTTCAGTT 109
Db 2 MetArgAlaArgPhePro-----LeuLeuLeuGlyLeuValPheLeuAlaSerVal 19
QY 110 TCTGCAACGCATGCCCAAGTCATCATCTTACCAGAGAAACAGAGAACCCCTGCGCCAG 169
Db 20 SerValSerPheGlyIleAla-----TyrTrpGluLys---GluAsnProLysHisAsn 36
QY 170 AGGTGCTCCAGAGTTGTCAACAGGAAACCGGATGACTTGAAGCAAAAGCATCGGAGTCT 229
Db 37 LysCysLeuGlnSerCysAsnSerGluArgAspSerTyrArgAsnGlnAlaCysHisAla 56
QY 230 CGCTGCACCAAGCTCGAGTATGAT-----CCTCGTTGT 262

```
Db 57 ArgCysAsnLeuLeuLysValGluLysGluGluCysGluGluGlyGluLeuProArg--- 75
||||| ||||| :|||
QY 263 GTCTATGATCCTCGAGGACACTGGCACCACCAACCAAGTTCCTCCAGGCGG 322
||||| ||||| |||||
Db 76 -----ProArgProArg-----ProGlnHisProGlu 84
||||| ||||| |||||
QY 323 ACAGTGGCCGCAACCCGAGACTAGATGATGAC----- 358
:||||| ||||| :|||||
Db 85 ArgGluProGlnProGlyGluLysGluGluAspGluGluProArgProIle 104
||||| ||||| |||||
QY 359 -----CGCGTCAACCCCAAGAGAG-----GAAGGAGCCGA 391
||||| ||||| |||||
Db 105 ProPheProArgProGlnProArgGlnGluGluHisGluGlnArgGluGlnGlu 124
||||| ||||| |||||
QY 392 TGG-----GGACCAGCTGGACCGGAGGAGCGGTGAAGAGAGAA 430
||||| ||||| |||||
Db 125 TrpProArgLysGluGluLysArgGlyGluLysGlySerGluGluAspGlu 144
||||| ||||| |||||
QY 431 GACTGGAGACAACAGAGAGAT-----TGGAGGCGCAACAGTCAACGACCCG 484
||||| ||||| |||||
Db 145 AspGluGlnAspGluArgGlnPheProPheProArgProHisGlnLysGlu 164
||||| ||||| |||||
QY 485 AAAATAAGGCCGCAAGAGAGAGAGAGAGAGTGGGGAACACACCGTAGCCATGTG 544
||||| ||||| |||||
Db 165 ArgAsnGluGluGluAspGluAspGluGlnGlnArgGluSerGluSerGluAsp 184
||||| ||||| |||||
QY 545 AGGGAA-----GAAACATCTCGAACAACCTTTCTACTTCCTCCGCAAGCGGTTAGC 598
||||| ||||| |||||
Db 185 SerGluLeuArgArgHisLysAsnLysAsnProPheLeuPheGlySerAsnArgPheGlu 204
||||| ||||| |||||
QY 599 ACCCGTACGGGAACCAACCGTAGATCCGGTCTCGAGAGGTTTGACCAAGGTCA 658
||||| ||||| |||||
Db 205 ThrLeuPheLysAsnGlnTyrGlyArgIleArgValLeuGlnArgPheAsnGlnArgSer 224
||||| ||||| |||||
QY 659 AGCAGTTTCAGATCTCCAGATCCAGTATGTCAGATCCAGGCCAACCTACACT 718
||||| ||||| |||||
Db 225 ProGlnLeuGlnAsnLeuArgAspTyrArgIleLeuGluPheAsnSerLysProAsnThr 244
||||| ||||| |||||
QY 719 CTGTGTTCTCCCAAGCAGCTGATGTGATAACATCTCTGTTATCCAGCAAGGCGCAAGCC 778
||||| ||||| |||||
Db 245 LeuLeuLeuProAsnHisAlaAspAlaAspTyrLeuIleValIleLeuAsnGlyThrAla 264
||||| ||||| |||||
QY 779 ACCGTGACCTAGCAATGCAATACAGAAAGAGCTTTAATCTTGACGAGGCGCATGCA 838
||||| ||||| |||||
Db 265 IleLeuSerLeuValAsnAsnAspArgAspSerTyrArgLeuGlnSerGlyAspAla 284
||||| ||||| |||||
QY 839 CTCAGATCCCATCCCGTTTCATTTCTCTACATCTTGAACCGCATGACACACAGACCTC 898
||||| ||||| |||||
Db 285 LeuArgValProSerGlyThrThrTyrTyrValValAsnProAspAsnAsnGluAsnLeu 304
||||| ||||| |||||
QY 899 AGATAGCTAAATCTCCATGCCGCTTAACACACCCGCGCAGTTTGAGGATTTCTCCCG 958
||||| ||||| |||||
Db 305 ArgLeuIleThrLeuAlaIleProValAsnLysProGlyArgPheGluSerPheLeu 324
||||| ||||| |||||
QY 959 GCGAGCAGCGGACCAATCATCTACTTCAGGGCTTCAGCAGGAATACGTTGGAGGCC 1018
||||| ||||| |||||
Db 325 SerSerThrGluAlaGlnSerTyrLeuGlnGlyPheSerArgAsnIleLeuGluAla 344
||||| ||||| |||||
QY 1019 GCCTTCAATCCGGAATCAATGATAGATACGAGGCTGCTGTAGNAGAGATGCGAGGTT 1078
||||| ||||| |||||
Db 345 SerTyrAspThrLysPheGluGluIleAsnLysValLeuPheSerArgGluGluGly--- 363
||||| ||||| |||||
QY 1079 GAGCAGAGGAGGAGGAGGAGGAGTGGAGTACTCGGAGTAGTGAGAACAAATGAAGGA 1138
||||| ||||| |||||
Db 364 -----GlnGlnGlnGluGlnArgLeu-----GlnGluSer 374
||||| ||||| |||||
QY 1139 GTGATAGTCAAGTGTCAAGGAGCACGTTGAAGAACTTACTAAGCACGCTAAATCCGTC 1198
||||| ||||| |||||
Db 375 ValIleValGluIleSerLysGluGlnIleArgAlaLeuSerLysArgAlaLysSerSer 394
||||| ||||| |||||
QY 1199 TCAAGAAAGGCTCCGAGAGAGGAGGAGATATCACCAACCAATCACTTGACGAGGCG 1258
||||| ||||| |||||
```

```
Db 395 SerArgLysThrIleSerSerGlu-----AspLysProPheAsnLeuArgSerArg 411
||||| ||||| |||||
QY 1259 GAGCCGATCTTTTCTAAACAACCTTTGGGAAGTTATTTGAGGTGAAGCCAGACAAGAAC 1318
:||||| ||||| |||||
Db 412 AspProIleTyrSerAsnLysLeuGlyLysPhePheGluIleThrProGlu---LysAsn 430
||||| ||||| |||||
QY 1319 CCCAGCTTCAGGACCTGACATGATGCTCCTCCTGTTAGATCAAGAGAGAGCTTTG 1378
||||| ||||| |||||
Db 431 ProGlnLeuArgAspLeuAspIlePheLeuSerIleValAspMetAsnGluGlyAlaLeu 450
||||| ||||| |||||
QY 1379 ATGCTCCCACTTCAACTCAAGGCGCATGGTTATCGTCTGCTCAACAAAGGAACCTGGA 1438
:||||| ||||| |||||
Db 451 LeuLeuProHisPheAsnSerLysAlaIleValIleLeuValIleAsnGluGlyAspAla 470
||||| ||||| |||||
QY 1439 AACCTTGAACCTGCTGTAAAGAAAGAGACAACAAGAGGCGGCGGGAAGAGAG 1498
||||| ||||| |||||
Db 471 AsnIleGluLeuValGlyLeu---LysGluGlnGlnGlnGlnGlnGluGln 489
||||| ||||| |||||
QY 1499 GAGGACAGACGACGAAGAGGAGGAGAGTACAGAGAGGTGCTAGGTACACAGCGAGG 1558
||||| ||||| |||||
Db 490 Pro-----LeuGluValArgLysTyrArgAlaGlu 499
||||| ||||| |||||
QY 1559 TTGAAGAAAGCGATGTTTCATGCTCCAGCAGCTCATCCAGTAGCCATCAACGCTTC 1618
||||| ||||| |||||
Db 500 LeuSerGluGlnAspIlePheValIleProAlaGlyTyrProValValValAsnAlaThr 519
||||| ||||| |||||
QY 1619 TCCGAACCTCATCTGCTTGGCTTCCGTTATCAACGCTGAAACCAACACAGAAATCTTCCTT 1678
||||| ||||| |||||
Db 520 SerAsnLeuAsnPhePheAlaIleGlyIleAsnAlaGluAsnAsnGlnArgAsnPheLeu 539
||||| ||||| |||||
QY 1679 GCAGGTGATAAGACAAATGTATAGACCATAGACAGACGACGAGGAGGATTTAGCATTC 1738
||||| ||||| |||||
Db 540 AlaGlySerGlnAspAsnValIleSerGlnIleProSerGlnValGlnGluLeuAlaPhe 559
||||| ||||| |||||
QY 1739 CCTGGGTCCGGTGAAACAAGTTGAGAAGCTCATCAAAACCCAGAAAGAAATCTCACTTTGTG 1798
||||| ||||| |||||
Db 560 ProGlySerAlaGlnAlaValGluLysLeuLeuLysAsnGlnArgGluSerTyrPheVal 579
||||| ||||| |||||
QY 1799 AGTGTCTGCTCAATCTCAATCTCAATCTCCGCTCTCTCTGAGAAAGAGTCTCCTGAG 1858
||||| ||||| |||||
Db 580 AspAlaGlnProLysLys----- 585
||||| ||||| |||||
QY 1859 AAAGAGGATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1918
:||||| ||||| |||||
Db 586 -----LysGluGluGlyAsnLysGlyArgLysGlyProLeuSerSerIleLeuArg 602
||||| ||||| |||||
QY 1919 GCTTTT 1924
||||| ||||| |||||
Db 603 AlaPhe 604
||||| ||||| |||||
```

Search completed: April 8, 2005, 06:48:50
Job time : 276.254 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 8, 2005, 04:11:27 ; Search time 54.9253 Seconds
(without alignments)
5523.384 Million cell updates/sec

Title: US-10-728-051-1

Perfect score: 3562

Sequence: 1 aataacatatattcatc.....cggttggtggtttcttc 2032

Scoring table:

BLOSUM62	Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Ygapext 0.5	
Xgapop 6.0 , Ygapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 1027090

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10728051/runat_07042005_125046_14072/app.query.fasta_1.4757
-DB=Issued Patents AA -CPM=fastan -SUFFIX=rai -MINMATCH=0 1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10728051@cgn 1 1 122 @runat_07042005_125046_14072 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3278	92.0	626	4	US-09-106-872A-4
2	1700	47.7	335	4	US-09-106-872A-17
3	1279	35.9	605	1	US-07-955-905A-24
4	1267	35.6	571	1	US-07-955-905A-25
5	1068	30.0	410	1	US-07-955-905A-26
6	811	22.8	566	1	US-07-955-905A-22
7	811	22.8	566	1	US-07-955-905A-22
8	788	22.1	421	1	US-07-955-905A-27
9	777	21.8	587	1	US-07-955-905A-23
10	591.5	16.6	489	4	US-09-424-283-3
11	569	16.0	524	4	US-09-424-283-1
12	559.5	15.7	448	4	US-09-323-195A-18

13	514.5	14.4	409	4	US-09-424-283-4	Sequence 4, Appli
14	513.5	14.4	523	4	US-09-323-195A-17	Sequence 17, Appl
15	501	14.1	444	4	US-09-424-283-2	Sequence 2, Appli
16	183	5.1	762	4	US-09-352-991A-29423	Sequence 29423, A
17	173	4.9	1564	4	US-10-144-198-2	Sequence 2, Appli
18	173	4.9	1564	4	US-10-144-198-4	Sequence 4, Appli
19	166.5	4.7	720	4	US-09-352-991A-21881	Sequence 21881, A
20	164.5	4.6	1266	4	US-09-352-991A-30851	Sequence 30851, A
21	163	4.6	714	4	US-09-949-016-10296	Sequence 10296, A
22	160	4.5	630	4	US-09-352-991A-26324	Sequence 26324, A
23	160	4.5	801	4	US-09-352-991A-29274	Sequence 29274, A
24	160	4.5	1898	1	US-08-056-200-94	Sequence 94, Appl
25	160	4.5	1898	2	US-08-056-200-94	Sequence 94, Appl
26	160	4.5	1898	4	US-09-538-092-1280	Sequence 1280, Ap
27	159	4.5	1043	4	US-09-352-991A-28885	Sequence 28885, A
28	158.5	4.4	1075	4	US-09-352-991A-18387	Sequence 18387, A
29	157	4.4	1780	1	US-08-769-309A-5	Sequence 5, Appli
30	157	4.4	1780	3	US-08-984-570-5	Sequence 5, Appli
31	157	4.4	1781	4	US-09-361-403-13	Sequence 13, Appl
32	157	4.4	2079	4	US-09-949-016-8301	Sequence 8301, Ap
33	155	4.4	754	4	US-09-976-594-375	Sequence 375, App
34	155	4.4	760	4	US-09-949-016-11129	Sequence 11129, A
35	155	4.4	863	4	US-09-352-991A-26099	Sequence 26099, A
36	154	4.3	650	4	US-09-352-991A-20375	Sequence 20375, A
37	153.5	4.3	1233	4	US-09-688-188B-89	Sequence 89, Appl
38	153.5	4.3	1233	4	US-09-291-417D-89	Sequence 89, Appl
39	152.5	4.3	580	4	US-09-348-796A-14473	Sequence 14473, A
40	152	4.3	1162	2	US-08-728-323A-2	Sequence 2, Appli
41	152	4.3	1162	3	US-09-298-568-2	Sequence 2, Appli
42	152	4.3	1162	4	US-09-410-399-2	Sequence 2, Appli
43	152	4.3	1162	4	US-09-894-273-2	Sequence 2, Appli
44	152	4.3	1427	4	US-09-252-991A-27005	Sequence 27005, A
45	151.5	4.3	917	4	US-09-252-991A-25101	Sequence 25101, A

ALIGNMENTS

RESULT 1

US-09-106-872A-4

; Sequence 4, Application US/09106872A

; Patent No. 6486311

; GENERAL INFORMATION:

; APPLICANT: Burks Jr., A. Wesley

; APPLICANT: Stanley, J. Steven

; APPLICANT: Cockrell, Gael

; APPLICANT: King, Nina E.

; APPLICANT: Sampson, Hugh A.

; APPLICANT: Helm, Ricki M.

; APPLICANT: Bannon, Gary A.

; TITLE OF INVENTION: Peanut Allergens and Methods

; FILE REFERENCE: HS 103 CIP

; CURRENT APPLICATION NUMBER: US/09/106,872A

; CURRENT FILING DATE: 1999-06-29

; PRIOR APPLICATION NUMBER: PCT/US96/15222

; PRIOR FILING DATE: 1996-09-23

; NUMBER OF SEQ ID NOS: 23

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 4

; LENGTH: 626

; TYPE: PRT

; ORGANISM: Arachis hypogaea

; FEATURE:

; OTHER INFORMATION: Amino Acids 25-34 are Ara H 1 binding epitope,

; OTHER INFORMATION: peptide 1

; OTHER INFORMATION: Amino Acids 48-57 are Ara H 1 binding epitope,

; OTHER INFORMATION: peptide 2

; OTHER INFORMATION: Amino Acids 65-74 are Ara H 1 binding epitope,

; OTHER INFORMATION: peptide 3

; OTHER INFORMATION: Amino Acids 89-98 are Ara H 1 binding epitope,

; OTHER INFORMATION: peptide 4

; OTHER INFORMATION: Amino Acids 97-106 are Ara H 1 binding epitope,

; OTHER INFORMATION: peptide 5

; OTHER INFORMATION: Amino Acids 107-116 are Ara H 1 binding epitope,

/	OTHER INFORMATION: peptide 6																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	</
---	------------------------------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----

121	ArgGluArgGluArgGluGluAspTrpArgGlnProArgGluAspTrpArgArgProSer	140
470	CATCAGCAGCCAGCAAAATAAGCCCGAAGAGAGAGAGCAAGAGTGGCGGAACA	529
141	HisGlnGlnProArgLysIleArgProGluGlyArgGluGlyGlnGlnTrpGlyThr	160
530	CCAGTAGCCATGTGAGGGAAGAAACATCTCGGAACAACCCCTTCTTACTTCCCGTCAAGG	589
161	ProGlySerHisValArgGluGluThrSerArgAsnAsnProPheTyrPheProSerArg	180
590	CGGTTTAGCACCCTGACGGNACCAAAACGGTAGGATCCGGTCTCCAGAGGTTTCAC	649
181	ArgPheSerThrArgTyrGlyAsnGlnAsnGlyArgIleArgValLeuGlnArgPheAsp	200
650	CAAAGGTCAAGGCAGTTTCAGAAATCTCCAGAAATCACCGTATTGTGCAGATCAGAGCCAAA	709
201	GlnArgSerArgGlnPheGlnAsnLeuGlnAsnHisArgIleValGlnIleGluAlaLys	220
710	CCTAACACTCTTGTCTTCCCAAGCACGCTGATGCTGATAACATCCTTGTGTATCCAGCA	769
221	ProAsnThrLeuValLeuProLysHisAlaAspAlaAspAsnIleLeuValIleGlnGln	240
770	GGGCAAGCACCGCTGACCGTACCAATGCAATACAGAAAGAGCTTTAATCTTCACGAG	829
241	GlyGlnAlaThrValThrValAlaAsnGlyAsnAsnArgLysSerPheAsnLeuAspGlu	260
830	GGCCATGCACCTCAGAAATCCCATCCGTTTTCATTTCTCTACATCTTGAACCGCATGACAAC	889
261	GlyHisAlaLeuArgIleProSerGlyPheIleSerTyrIleLeuAsnArgHisAspAsn	280
890	CAGAACCTCAGAGTAGCTAAATCTCCATGCCCGCTTAACACACCCGGCCAGTTTCCAGAT	949
281	GlnAsnLeuArgValAlaLysIleSerMetProValAsnThrProGlyGlnPheGluAsp	300
950	TTCTTCCCGCGCAGCAGCCGACCAATCATCTACTTGCAGGGCTTCAGCAGGGAATACG	1009
301	PhePheProAlaSerSerArgAspGlnSerSerTyrLeuGlnGluPheSerArgAsnThr	320
1010	TTGGAGGCGCGCTTCAATCGCGAATTCATAGATACGAGGGTCTCTGTAGACAGAAAT	1069
321	LeuGluAlaAlaPheAsnAlaGluPheAsnGluIleArgArgValLeuLeuGluAsn	340
1070	GCAGAGGTGACCAAGAGAGAGAGCGGACGCGCATCGGAGTACTCGGAGTAGTCAGAC	1129
341	AlaGlyGlyGluGlnGluArgGlyGlnArgArgTrpSerThrArgSerSerGluAsn	360
1130	AATGAAGGAGTGCATAGTCAAAGTGTCAAAGGAGCAGCTTGAAGAACTTACTTAAGCACGCT	1189
361	AsnGluGlyValIleValLysValSerLysGluHisValGluGluLeuThrLysHisAla	380
1190	AAATCCGCTCTCAAGAAAGGCTCCGAAAGAGGAGATATCAACCAACCAATCAACTTG	1249
381	LysSerValSerLysLysGlySerGluGluGluGlyAspIleThrAsnProIleAsnLeu	400
1250	AGAGAAGGCGACCCGATCTTCTAACACTTGGAGTATTTTTCAGGTGAGGCGCAGAC	1309
401	ArgGluGlyGluProAspLeuSerAsnAsnPheGlyLysLeuPheGluValLysProAsp	420
1310	AAGAAGAACCCCGAGCTTCAGGACCTGGACATGATGCTCACCTGTGTGAGAGATCAAGAA	1369
421	LysLysAsnProGlnLeuGlnAspLeuAspMetMetLeuThrCysValGluIleLysGlu	440
1370	GGAGCTTTGATGCTCCCACTTCAACTCAAGGCCATGGTTATCTGCTCGTCAACAA	1429
441	GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValValAsnLys	460
1430	GGAACTGGAAACCTTGAACCTCGTGGCTGTAAGAAAGAGCAACACAGAGGCGACGCGG	1489
461	GlyThrGlyAsnLeuGluLeuValAlaValArgLysGluGlnGlnArgGlyArgArg	480
1490	GAAAGAGAGGAGGACCAAGACCAAGAGAGGAGGGAAGTAACAGAGAGTGGTGGTAC	1549
481	GluGluGluGluAspGluGluGluGluGlySerAsnArgGluValArgArgTyr	500

Qy	1550	ACACGAGGTTGAAGGAAGCGGATGTGTTCTATCATGCGCAGCAGCTCATCCAGTAGCCATC	1609
Db	501	ThrAlaArgLeuLysGluGlyAspValPheIleMetProAlaAlaHisProValAlaIle	520
Qy	1610	AACGCTTCCTCCGAACCTCCATCTCTGCTGGCTTCGGTATCAACGCTGAAAAACACCAACAGA	1669
Db	521	AsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsnAsnHisArg	540
Qy	1670	ATCTTCCTTCGAGTGATAAGGACAAATGTGATACACCAGATAGAGAAGCAAGCGAAGCAT	1729
Db	541	IlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGlnAlaLysAsp	560
Qy	1730	TTAGCATTCCTCGGTCGGGTGAACAGTTTGAGAAGCTCATCAAAAACCAAGAGGAATCT	1789
Db	561	LeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGlnLysGluSer	580
Qy	1790	CACTTTCTGAGTGCTCGCTCCTCAATCTCAATCTCAATCTCCGTCGTCTCCTGAGAAAGAG	1849
Db	581	HisPheValSerAlaArgProGlnSerGlnSerGlnSerProSerSerProGluLysGlu	600
Qy	1850	TCCTCTGAGAAAGAGGATCAAGAGGAGGAAAAACCAAGAGGAGGAGGTCCACCTCTTTTCA	1909
Db	601	SerProGluLysGluAspGlnGluGluAsnGlnGlyGlyLysGlyProLeuLeuSer	620
Qy	1910	ATTTTGAAGCCTTTTAAAC	1927
Db	621	IleLeuLysAlaPheAsn	626

RESULT 2

US-09-106-872A-17
; Sequence 17, Application US/09106872A

; Patent No. 6486311

```

/ GENERAL INFORMATION:
/ APPLICANT: Burks Jr., A. Wesley
/ APPLICANT: Stanley, J. Steven
/ APPLICANT: Cockrell, Gael
/ APPLICANT: King, Nina E.
/ APPLICANT: Sampson, Hugh A.
/ APPLICANT: Helm, Ricki M.
/ APPLICANT: Bannon, Gary A.
/ TITLE OF INVENTION: Peanut Allergens and Methods
/ FILE REFERENCE: HS 103 CIP
/ CURRENT APPLICATION NUMBER: US/09/106.872A
/ CURRENT FILING DATE: 1999-06-29
/ PRIOR APPLICATION NUMBER: PCT/US96/15222
/ PRIOR FILING DATE: 1996-09-23
/ NUMBER OF SEQ ID NOS: 23
/ SOFTWARE: Patent In Ver. 2.1

```

Alignment Scores:					
Pred. No.:	8,73e-156	Length:	335		
Score:	1700.00	Matches:	335		
Percent Similarity:	99.41%	Conservative:	0		
Best Local Similarity:	99.41%	Mismatches:	0		
Query Match:	47.73%	Indels:	2		
DB:	4	Gaps:	1		

US-10-728-051-1 (1-2032) X US-09-106-872A-17 (1-335)

917	ATGCCCGTTTAA	CACACCCGCGCAGTTT	TGGAGATTTCTTCCCGCGAGCAGCCGAGACAA	976
Qy				
1	MetProValAsnThr	ProGlyGlnPheGlu	AspPheProIaSerSerArgAspGln	20
Db				
977	TCATCTTACTTGC	AGGCGCTTCAGCAGAAATACGTTGGAGGCGCGCTTCAATCGGGGAATC	1036	
Qy				
21	SerSerTyrLeuGlnGlyPhe	SerArgAsnThrLeuGluIaIaLapheAsnAlaGluPhe	40	
Db				

Qy	1037	AATGATAGATCGGAGGTGCTGTTAGAGAGAAATCAGGAGGTGAGCAAGAGGAGAGAGGG	1098
Db	41	AsnGluIleargValLeuLeuGluGluAsnAlaGlyGlyGluGluGluArgGly	60
Qy	1097	CAGAGGCGATCGAGTACTCCGAGTAGTGAGAACAAATCAAGGAGGTAGTAGTCAAAAGTGTCA	1156
Db	61	GlnArgArgIrpSerThrArgSerSerGluAsnAsnGluGlyValIleValSer	80
Qy	1157	AAGGAGCACGTTGAAGAAGCTTACTTAAGCAGCGCTAAATCCGTCTCAAGAAAGGCTCCGAA	1216
Db	81	LysGluHisValGluGluLeuThrLysHisAlaLysSerValSerLysGlySerGlu	100
Qy	1217	GAAGAGGAGATATCACCAACCCAAATCAACTGAGAGAAAGCGAGCCCGATCTTTCTTAAC	1276
Db	101	GluGluGlyAspIleThrAsnProIleAsnLeuArgGluGlyGluProAspLeuSerAsn	120
Qy	1277	AACTTTGGGGAAGTTATTTGAGGTGAAGCCAGACAAAGAAACCCCAAGCTTCAGGACCTG	1336
Db	121	AsnPheGlyLysLeuPheGluValLysProAspLysLysAsnProGlnLeuGlnAspLeu	140
Qy	1337	GACATGATGCTCACCTGTGTAGAGATCAAAAGAAAGGAGCTTTGTAGTCTCCACACTTCAAC	1396
Db	141	AspMetMetLeuThrCysValGluIleLysGluGlyAlaLeuMetLeuProHisPheAsn	160
Qy	1397	TCAAGGCCCATGTTATTCGTGCTGCTCAACAAAGAAAGAACTGGAACTCTGCTGGCT	1456
Db	161	SerLysAlaMetValIleValValValAsnLysGlyThrGlyAsnLeuGluLeuValAla	180
Qy	1457	GTAAGAAAGAGCAACACAGAGGGGACGCGGAGAGAGAGGAGGAGGAGGAGGAGAA	1516
Db	181	ValArgLysGluGlnGlnArgIleArgLysGluGluAspGluGluAspGluGlu	200
Qy	1517	GAGGAGGGAAGTAACAGAGAGGTGCGTAGGTACACAGCGAGGTTGAAGGAAGCGGATGTG	1576
Db	201	GluGluGlySerAsnArgGluValArgArgTyrThrAlaArgLeuLysGluGlyAspVal	220
Qy	1577	TTCATCATGCCAGCGCTCATCCAGTAGCCATCAACGCTTCTCCGAACTCCATCTGCTT	1636
Db	221	PheIleMetProAlaAlaHisProValAlaIleAsnAlaSerSerGluLeuHisLeuLeu	240
Qy	1637	GGCTTCGGTATCAACCTGAAACCAACCACAGAAATCTTCTTCAGGTGATAGGACAAT	1696
Db	241	GlyPheGlyIleAsnAlaGluAsnAsnHisArgIlePheLeuAlaGlyAspLysAspAsn	260
Qy	1697	GTGATAGCAGATAGAGAAGCAAGCAAGGATTTAGCATTTCCCTGGTCCGGTGAACAA	1756
Db	261	ValIleAspGlnIleGluLysGlnAlaLysAspLeuAlaPheProGlySerGlyGluGln	280
Qy	1757	GTTGAGAAGCTCATCAAAACCAGAAAGGAATCTCACTTGTGTAGTGTCTGCTCTCAATCT	1816
Db	281	ValGluLysLeuIleLysAsnGlnLysGluSerHisPheValSerAla-----GlnSer	298
Qy	1817	CAATCTCAATCTCGTGTCTCTGAGAAAGAGTCTCTGAGAAAGAGGATCAAGAGGAG	1876
Db	299	GlnSerGlnSerProSerProGluLysGluSerProGluLysGluAspGlnGluGlu	318
Qy	1877	GAATAACCAAGGAGGAGGTCCACCTCTTCAATTTTGAAGGCTTTTAAC	1927
Db	319	GluAsnGlnGlyLysGlyProLeuLeuSerIleLeuLysAlaPheAsn	335
RESULT 3			
US-07-955-905A-24			
; Sequence 24, Application US/079555905A			
; Patent No. 5770433			
; GENERAL INFORMATION:			
; APPLICANT:			
; TITLE OF INVENTION: RECOMBINANT 47 AND 31 kD COCOA PROTEINS AND			
; PRECURSOR			
; NUMBER OF SEQUENCES: 28			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			

Db 540 AladlySerGlnAepAenValIleSerGlnIleProSerGlnValGlnGluLeuAlaPhe 559
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Qy 1739 CCTGGGTCGGGTGAACAAAGTTGAGAGCTCATCAAAAACAGAGGAATCTCACTTTGTG 1798
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Db 560 ProGlySerAlaGlnAlaValGluLyLeuLyLeuLyLeuLyLeuLyLeuLyLeuLyLeu 579
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Qy 1799 AGTCTCGTCTCAATCTCAATCTCAATCTCGCTCGTCTCTCGAGAAAGAGTCTCTGAG 1858
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Db 580 AspAlaGlnProLyLyLyLyLyLyLyLyLyLyLyLyLyLyLyLyLyLyLyLyLyLy 585
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Qy 1859 AAAGAGGATCAAGAGGAGGAGAAACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1918
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Db 586 -----LysGluGluGluGlyAsnLysGlyArgLysGlyProLeuSerSerIleLeuArg 602
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Qy 1919 GCTTTT 1924
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Db 603 AlaPhe 604

RESULT 4

US-07-955-905A-25
; Sequence 25, Application US/07955905A
; Patent No. 5770433
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: RECOMBINANT 47 AND 31 kD COCOA PROTEINS AND
; TITLE OF INVENTION: PRECURSOR
; NUMBER OF SEQUENCES: 28
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/955,905A
; FILING DATE: 21-JAN-1993
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 571 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORGANISM: Pisum sativum
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..571
; OTHER INFORMATION: /note= "Convicillin from P. sativum"
US-07-955-905A-25

Alignment Scores:
Pred. No.: 1.13e-113 Length: 571
Score: 1267.00 Matches: 276
Percent Similarity: 56.67% Conservative: 98
Best Local Similarity: 41.82% Mismatches: 146
Query Match: 35.57% Indels: 140
DB: 1 Gaps: 13

US-10-728-051-1 (1-2032) x US-07-955-905A-25 (1-571)

Qy 68 CCACTGATGCTGTGTAGGATCTTGTCTCTGCTTCAGTTCTGCAACGCATGCCAAG 127
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Db 10 ProLeuLeuLeuPheLeuGlyIlePheLeuAlaSerValCysValThrTyAlaAsn 29
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Qy 128 TCATCCTTACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 187
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Db 29 ----- 29
Qy 188 CAACAGGAACCGATGACTTGAAGCAAAAGGATGCGAGTCTCGCTGCACCAAGCTCGAG 247
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Db 29 ----- 29

Qy 248 TATGATCTCGTGTGTCTATGATCTCTGAGGACACACTGGCACCACCAACCAACGTTCC 307
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Db 30 -----TyrAspGluGlySerGluThrArgVal----- 38
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Qy 308 CTTCCAGGGAGCGGACACCTGCGCCCAACCCGAGACTTACCATGATGATGATGATGATGAT 367
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Db 39 ---ProGlyGlnArgGluArgGlyArgGlnGluGlyGlyGlyGlyGlyGlyGlyGlyGly 56
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Qy 368 CCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 418
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Db 57 -----GlyGluTrpArgProSerTyrGluLysGluGluHisGluGlu 70
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Qy 418 ----- 418
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Db 71 GluLysGlnLysTyrArgTyrGlnArgGluLysGluGlnLysGluValGlnProGly 90
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Qy 419 -----GAAAGAGAAGAGACTGGAGACAACCAAGACAAGATTCGAGG----- 460
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Db 91 ArgGluArgTrpGluArgGluGluAspGluGluValGluGluTrpArgGlySer 110
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Qy 461 -----CGACCAAGTCTATCAGCAGCCACCG 484
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Db 111 GlnArgArgGluAspProGluGluArgAlaArgLeuArgHisArgGluGluArgThrLys 130
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Qy 485 AAAATAAGCCCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 544
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Db 131 ArgAspArgArgHisGlnArgGluGlyGluGlu-----GluArgSer 145
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Qy 545 AGGGAAGAAACATCTCGGAACACCTTTCTACTTCCGTCAGAGCGGTTTACACCCGC 604
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Db 146 SerGluSerGlnGluHisArgAsnProPheLeuPheLysSerAsnLysPheLeuThrLeu 165
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Qy 605 TACGGGAACCAAAACGGTAGGATCCGGGTCTGTCAGAGGTTTACCAAAAGTCAAGGCAG 664
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Db 166 PheGluAsnGluAsnGlyHisIleArgArgLeuGlnArgPheAspLysArgSerAspLeu 185
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Qy 665 TTTCAGAAATCTCAGAAATCACGTATTTGTCAGATCGAGGCCAAACCTTAACATCTTGT 724
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Db 186 PheGluAsnLeuGlnAsnTyrArgLeuValGluTyrArgAlaLysProHisThrIlePhe 205
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Qy 725 CTTCCCAAGCAGCGTCTGATGCTGATTAACATCTTGTATCCAGCAAGGCAAGCCAGCGT 784
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Db 206 LeuProGlnHisIleAspAlaAspLeuIleLeuValLeuAsnGlyLysAlaIleLeu 225
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Qy 785 ACCGTAGCAAAATGGCAATAACAGAAAGAGCTTTAATCTTTCAGCAGGCGCATGCATCAGA 844
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Db 226 ThrValLeuSerProAsnAspArgAsnSerTyrAsnLeuGluArgGlyAspThrIleLys 245
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Qy 845 ATCCCATCCGGTTCATTTCTTACATCTTGAACCCGCATGACAAACAGAACCTCAGAGTA 904
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Db 246 IleProAlaGlyThrThrSerTyrLeuValAsnGlnAspGluGluAspLeuArgVal 265
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Qy 905 GCTAAAATCTCCATGCCGTTAACACACCCGCGCAGTTTGAGGATTTCTCCCGCGCAGC 964
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Db 266 ValAspPheValIleProValAsnArgProGlyLysPheGluAlaPhe-----GlyLeu 283
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Qy 965 AGCCGAGACCAATCATCTTCTTTCAGAGGCTTTCAGCAGGAATACGTTGGAGCGCGCTTC 1024
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Db 284 SerGluAsnLysAsnGlnTyrLeuArgGlyPheSerLysAsnIleLeuGluAlaSerLeu 303
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Qy 1025 AATGCGGAATTCATGAGATACGAGGCTGCTGTTAGAGAGAGATGAGAGAGTGAACAA 1084
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Db 304 AsnThrLysTyrGluThrIleGluLysValLeuGluGluGluGluGluGluGluGluGlu 323
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Qy 1085 GAGGAGAGGCGGAGAGGCGATGAGTACTCGGAGTAGTGAGAAACATGAAGGAGTGATA 1144
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Db 324 GlnLeuArgAspArgLysArg-----ThrGlnGlnGlyGluGluArgAsp-----AlaIle 340
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Qy 1145 GTCAAAAGTGTCAAGGAGCAGCTTGAAGAACTTTACTAAGCAGCGTAAATCCGCTCAAAG 1204
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Db 341 IleLysValSerArgGluGlnIleGluGluLeuArgLysLeuAlaLysSerSerLys 360
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |
Qy 1205 AAAGGCTCCGAGAGAGGAGAGATATCAACCAACCCCAATCACTTGAGAGAGGCGGCC 1264
| | | | | : : : | | | | | | | | | | | | | | | : : : : : | | | | |

```
Db 361 LysSerLeuProSerGluPheGlu-----ProPheAsnLeuArgSerHisLysPro 377
QY 1265 GATCTTTTAAACAACTTTGGGAAGTTATTGAGGTGAAGCCAGACAGAAAGAACCCCGAG 1324
Db 378 GluTySerAsnLysPheGlyLysLeuPheGluLeuThrProGluLysLysTyProGln 397
QY 1325 CTTGAGGACCTGACATGATGCTCACTGTGTAGATCAAGAGGAGCTTTGATGCTC 1384
Db 398 LeuGlnAspLeuAspLeuValSerCysValGluLeuAsnLysGlyAlaLeuMetLeu 417
QY 1385 CCACACTTCAACTCAAGGCGCATGTTATCGTGTGCTCAACAAAGAACTGGAACCTT 1444
Db 418 ProHisTyRanSerArgAlaLeuValLeuLeuValAsnGluGlyLysGlyAsnLeu 437
QY 1445 GAATCTGTGCTGTGAAGAAAGAGCAACAACAGAGGGCGCGGGAAGAGGAGGAC 1504
Db 438 GluLeuLeuGlyLysAsnGluGlnGlnGluArg-----449
QY 1505 GAAGACGAAGAGAGGAGGAGGAGTAAACAGAGAGTGCCTAGGTACACAGCGAGGTTGAAG 1564
Db 450 GluAspArgLysGluArg-----AsnAsnGluValGlnArgTyRArgLysGluAlaArgLeuSer 467
QY 1565 GAAGCGCATGTTTCATCATGCCAGCAGCTCATCCAGTACCCATCAACCGCTTCTCCGAA 1624
Db 468 ProGlyAspValValIleProAlaGlyHisProValAlaIleSerAlaSerAsn 487
QY 1625 CTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1684
Db 488 LeuAsnLeuLeuGlyPheGlyIleAsnAlaLysAsnGlnArgAsnPheLeuSerGly 507
QY 1685 GATAAGGACAATGTATAGACAGATAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1744
Db 508 SerAspAspAsnValIleSerGlnIleGluAsnProValLysGluLeuThrPheProGly 527
QY 1745 TCGGGTGAACAAAGTTGAGAAGCTCATCAAAACAGAGAAATCTCATCTTTGAGTGCT 1804
Db 528 SerSerGlnGluValAsnArgLeuIleLysAsnGlnLysGlnSerHisPheAlaSerAla 547
QY 1805 CGTCTCAATCTCAATCTCAATCTCCGCTCGTCTCTCGAGAAAGAGTCTCTCGAGAAAGAG 1864
Db 548 GluPro-----Glu 550
QY 1865 GATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1924
Db 551 GlnLysGluGluGluSerGlnArgLysArgSerProLeuSerSerValLeuAspSerPhe 570
```

RESULT 5

```
US-07-955-905A-26
; Sequence 26, Application US/07955905A
; Patent No. 5770433
;
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: RECOMBINANT 47 AND 31 kD COCOA PROTEINS AND
; TITLE OF INVENTION: PRECURSOR
; NUMBER OF SEQUENCES: 28
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/955,905A
; FILING DATE: 21-JAN-1993
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 410 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Pisum sativum
```

```
;
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..410
; OTHER INFORMATION: /note= "Vicilin from P. sativum"
US-07-955-905A-26
```

```
Alignment Scores:
Pred. No.: 1,77e-94 Length: 410
Score: 1068.00 Matches: 209
Percent Similarity: 71.64% Conservative: 84
Best local Similarity: 51.10% Mismatches: 104
Query Match: 29.98% Indels: 12
DB: 1 Gaps: 5
```

US-10-728-051-1 (1-2032) x US-07-955-905A-26 (1-410)

```
QY 542 GTGAGGGAAGAAACATCTCGGAACAACCCCTTTCTACTTCCCGTCAAGGCGGTTTAGCACC 601
Db 13 ValSerSerArgSerAspGlnGluAsnProPheIlePheLysSerAsnArgPheGlnThr 32
QY 602 CGCTACGGGAACCAAAACGGGTAGGATCCGGTCTCTGAGAGGTTTGACCAAGGTCAGG 661
Db 33 LeuTyRLeuAsnGluAsnGlyHisIleArgLeuLeuGlnLysPheAspLysArgSerLys 52
QY 662 CAGTTTCAAGATCTCCAGAATCACCGTATTGTCGAGATCGAGGCCAAACCTTAACACTCTT 721
Db 53 IlePheGluAsnLeuGlnAsnTyRArgLeuLeuGlnTyRLeuLysSerLysProHisThrLeu 72
QY 722 GTTCTTCCCAAGCAGCTGATGCTGATAACAATCTCTGTTATCCAGCAAGGCCAAGCCACC 781
Db 73 PheLeuProGlnTyRThrAspAlaAspPheIleLeuValValLeuSerGlyLysAlaThr 92
QY 782 GTGACCGTAGCAATGGCAATAACAGAAAGACTTTAATCTTGACGAGGCGCATGCACCTC 841
Db 93 LeuThrValLeuLysSerAsnAspAsnSerPheAsnLeuGluArgGlyAspAlaIle 112
QY 842 AGAATCCCATCCGTTTTCATTTCTTACATCTTGAACCGCCATGACAAACAGAACCTCAGA 901
Db 113 LysLeuProAlaGlySerIleAlaTyRPhelaAsnArgAspAsnGluProArg 132
QY 902 GTAGCTAAATATCCATCCCGTTAACACACCGGCCAGTTTGAGGATTTCTCCCGCGG 961
Db 133 ValLeuAspLeuAlaIleProValAsnLysProGlyGlnLeuGlnSerPheLeuLeuSer 152
QY 962 AGCAGCCGAGACCAATCATCTTACTTGACGGCTTTCAGAGGAATACGTTGGAGCGCCGCC 1021
Db 153 GlyThrGlnAsnGlnLysSerSerLeuSerGlyPheSerLysAsnIleLeuGluAla 172
QY 1022 TTCAATGCGGAATTCATAGATACGAGGTGCTGTTTAGAAGAGAAATCGAGGAGGTGAG 1081
Db 173 PheAsnThrAsnTyRLeuGluIleGluLysValLeuLeuGluGlnGln-----Glu 189
QY 1082 CAAGAGGAGAGGCGAGGCGATGGAGTACTCGAGTAGTGAG---AACAATGAAGA 1138
Db 190 GlnGluProGlnHisArgArgSerLeuLysAspArgGlnGlnIleAsnGluLeuAsn 209
QY 1139 GTGATAGTCAAAAGTGTCAAGGAGCAGCTGTGAAGAACTTACTTAAGCAGCTAAATCCGTC 1198
Db 210 ValIleValLysValSerArgAspGlnIleGluLeuSerLysAsnAlaLysSerSer 229
QY 1199 TCAAAGAAAGGCTCCGGAAGAGGAGGAGATATACCAACCCCAATCACTTGTAGAGAGGC 1258
Db 230 SerLysLysSerValSerSerGlu-----SerGlyProPheAsnLeuArgSerArg 246
QY 1259 GAGCCCGCATCTTCTAAACAACCTTGGGAAGTATTATTGAGTGAAGGCACACAGAAAGAAC 1318
Db 247 AsnProIleTyRSerAsnLysPheGlyLysPheGluIleThrProGlu---LysAsn 265
QY 1319 CCCAGCTTCAGGACCTGACATGATGCTCACCTGTGTAGAGATCAAGAGAGGCTTTG 1378
Db 266 GlnGlnLeuGlnAspLeuAspIlePheValAsnSerValAspIleLysValGlySerLeu 285
QY 1379 ATGCTCCCACTTCAACTCAAGGCGCATGTTATCGTCTCGTCAACAAAGGAACCTGGA 1438
```

```
Db 286 LeuLeuProAsnTyrAsnSerArgAlaIleValIleValThrValThrGluGlyLysGly 305
Qy 1439 AACCTTGAACCTCGTGGCTGTAAAGAAAGCAACAAACAGAGGGGACCGCGGAAGAGAG 1498
Db 306 AspPheGluLeuValGlyGlnArgAsnGluAsnGlnGlyLys-----GluAsn 321
Qy 1499 GAGGACGAAGACGAAGAGGAGGAGGAGTAACAGAGAGGTGCGTAGGTACACAGCGAGG 1558
Db 322 AspLysGluGluGlnGluGluGluThrSerLysGlnValGlnLeuTyrArgAlaLys 341
Qy 1559 TTGAAGGAAGCGGATGTTTCATCATCCAGCAGCTCATCCAGTAGCCATCAACGCTTCC 1618
Db 342 LeuSerProGlyAspValPheValIleProAlaGlyHisProValAlaIleAsnAlaSer 361
Qy 1619 TCCGAATCCATCTCGTTCGGTATCAACGCTCAAAACCAACCAACAGAAATCTTCCTT 1678
Db 362 SerAspLeuAsnLeuIleGlyLeuGlyLeuAsnAlaGluAsnGluArgAsnPheLeu 381
Qy 1679 GCAGGTGATAGGCAATGTGATAGACAGATAGAGCAAGCAAGGAGGATTTAGCATTC 1738
Db 382 AlaGlyGluGluAspAsnValIleSerGlnValGluArgProValLysGluLeuAlaPhe 401
Qy 1739 CCTGGTCCGGTGAACAAAGTTGAGAAG 1765
Db 402 ProGlySerSerHisGluValAspArg 410
```

RESULT 6

US-07-955-905A-2

; Sequence 2, Application US/07955905A

; Patent No. 5770433

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: RECOMBINANT 47 AND 31 kD COCOA PROTEINS AND

; TITLE OF INVENTION: PRECURSOR

; NUMBER OF SEQUENCES: 28

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/955,905A

; FILING DATE: 21-JAN-1993

; CLASSIFICATION: 435

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 566 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-07-955-905A-2

Alignment Scores:

```
Pred. No.: 1.75e-69 Length: 566
Score: 811.00 Matches: 195
Percent Similarity: 50.95% Conservative: 127
Best Local Similarity: 30.85% Mismatches: 218
Query Match: 22.77% Indels: 92
DB: 1 Gaps: 19
```

US-10-728-051-1 (1-2032) x US-07-955-905A-2 (1-566)

```
Qy 65 TCTCCATGATCGTGTGTAGGATCCTTGTCTCGGCTTCAGTTTCTGCAACGATGCC 124
Db 6 SerProPheIleValIlePheSerIleLeuLeuSerPheAlaLeuLeuCysSerGly 25
Qy 125 AAGTCATCATCTTACCAGAAAGAAACAGAACCCCTGCGCCAC-----AGGTGC 175
Db 26 ValSerAlaTyrGlyArgLysGlnTyrGluArgAspProArgGlnGlnTyrGluGlnCys 45
Qy 176 CTCAGAGTGTCAACAGGA-----CCGATGACTTGAAGCAAAAGGATCGAGTCTCGC 232
Db 333
```

```
Db 46 GlnArgArgCysGluSerGluAlaThrGluGluArgGluGlnGlnCysGluGlnArg 65
Qy 233 TGCACAAAGCTCAGATATGATCTCGTTGTGTCTATGATCTCGAGGACACACTGGCACC 292
Db 66 CysGluArg---GlnTyrLysGluGln----- 73
Qy 293 ACCAACCAAGTTCCTCCCTCCAGGGGAGCGGACACAGTGGCGCGCCCAACCCGGAGACTACGAT 352
Db 74 -----GlnArgGlnGlnGluGluGluLeuGlnArgGlnTyrGln----- 86
Qy 353 GATGACCGCGCTCAACCCCGAAGAGAGGAGGCGGATGGGGACCGAGCTGGACCGAGG 412
Db 87 -----GlnCysGlnGlyArgCysGlnGlnGlnGlnGln-----GlyGlnArg 100
Qy 413 GAGCGTGAA-----AGAGAAGAAGACTTGAGACAACCAAGAGAAGATTGAGG----- 460
Db 101 GluGlnGlnCysGlnArgLysCysTyrGluGlnTyrLysGluGlnArgGlyGlu 120
Qy 461 CGACCAAGTCATCAGCAGCCACCGAAATAAGGCCCGAAGAGAGAGAGAGAGAACAGAG 520
Db 121 HisGluAsnTyrHisAsnHisLysLysAsnArgSerGluGluGluGlnGlnGln--- 139
Qy 521 TGGGGAACACCAAGTAGCCATGTGAGGAAGAAACATCTCGGAACAACCTTTCTACTTC 580
Db 140 -----ArgAsnAsnProTyrTyrPhe 146
Qy 581 CCGTCAAGGCGG---TTTAGCACCCGCTACGGGAACCAAAACCGTAGGATCCGGGCTCTG 637
Db 147 ProLysArgArgSerPheGlnThrArgPheArgAspGluGluGluAsnPheLysIleLeu 166
Qy 638 CAGAGGTTTGACCAAAAGGTCAAGCAGTTTCAGAAATCTCCAGAAATCACCGTATTGTGCAG 697
Db 167 GlnArgPheAlaGluAsnSerProProLeuLysGlyIleAsnAspTyrArgLeuAlaMet 186
Qy 698 ATCGAGGCCAAACCTAACACTCTTGTCTTCCCAAGCAGCTCATCTGTATACATCCTT 757
Db 187 PheGluAlaAsnProAsnThrPheIleLeuProHisCysAspAlaGluAlaIleTyr 206
Qy 758 GTTATCCAGCAGGCAAGCCACCGTAGCAATGCGCAATAGCAATTAACAGAAAGAGCTTT 817
Db 207 PheValThrAsnGlyLysGlyThrIleThrPheValThrHisGluAsnLysGluSerTyr 226
Qy 818 AATCTTGACGAGGCGCATCAGATCCCATCCCGTTTCATTTCTCTATCATCTTGAAC 877
Db 227 AsnValGlnArgGlyThrValValSerValProAlaGlySerThrValTyrValValSer 246
Qy 878 CGCCATGACAAACCAAGACCTCAGAGTAGCTTAAATCTCCATGCGCGTTCACACACCGCGC 937
Db 247 GlnAspAsnGlnGluLysLeuThrIleAlaValLeuAlaLeuProValAsnSerProGly 266
Qy 938 CAGTTTGAGGATTTCTTCCCGCGCAGCAGCCGAGACCAATCATCTACTTTCAGGGGCTC 997
Db 267 LysTyrGluLeuPhePheProAlaGlyAsnAsnLysProGluSerTyrTyrGlyAlaPhe 286
Qy 998 AGCAGGAATACGTTGGAGGCGCTTCAATCGGAATTCATAGATACGAGAGGCTGCTG 1057
Db 287 SerTyrGluValLeuGluThrValPheAsnThrGlnArgGluLysLeuGluIleLeu 306
Qy 1058 TTAGAAGAGAATCCAGGAGGTGAGCAGAGGAGGCGCAGAGCGGATGAGTACTCGG 1117
Db 307 -----GluGluGlnArgGlyGlnLysArg----- 314
Qy 1118 AGTAGTGAGAAACAATGAAGAGGTGATAGTCAAAAGTGTCAAGAGGACGCTTGAGAACTT 1177
Db 315 ---GlnGlnGlnGlnGlnGlnMetPheArgLysAlaLysProGluGlnIleArgAlaIle 333
Qy 1178 ACTAAGCAGCGTAAATCTGCTCAAAAGAAAGGCTCCGAGAGAGAGGAGATATCACCAAC 1237
Db 334 SerGlnGlnAlaThrSerProArgHisArgGlyGlyGluArgLeuAla----- 349
Qy 1238 CCAATCAACTTGAGAGAGCGGAGCCGATCTTTCTTAACAACATTTTGGGAAGTATTGAG 1297
Db 350 ---IleAsnLeuLeuSerGlnSerProValTyrSerAsnGlnAsnGlnArgPheGlu 368
```

```
QY 1298 GTGAACCCACAGAGAACCCCGCTTTCAGACCTGAGATGATCTCACTGTGTA 1357
Db 369 AlaCysProGluAspPheSer---GlnPheGlnAsnMetAspValAlaValSerAlaPhe 387
QY 1358 GAGATCAAAAGAGAGCTTGTATGCTCCACACACTTCAACTCAAGGCCATGTTATCGTC 1417
Db 388 LysLeuAsnGlnGlyAlaIlePheValProHisTyrAsnSerLysAlaThrPheValVal 407
QY 1418 GTCGTCAACAAAGAACTGGAACCTTGAACCTCGTGTGCTGTAGAAAAGAGCAACAA--- 1474
Db 408 PheValThrAspGlyTyrGlyTyrAlaGlnMetAlaCysProHisLeuSerArgGlnSer 427
QY 1475 -----CAGAGGGAGCGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGA 1525
Db 428 GlnGlySerGlnSerGlyArgGlnAspArgArgGlnGlnGluGlnGluGlnGluGlu 447
QY 1526 AGTAACAGAGAGGTGCTAGGTACACAGCGAGGTGTAAGAGAGCGATGTTTCATCATG 1585
Db 448 ThrPheGlyGluPheGlnGlnValLysAlaProLeuSerProGlyAspValPheValAla 467
QY 1586 CCAGCAGCTCATCAGTAGCCATCAAGCTTCTCTCCGAA-----CTCCATCTGCTTGGC 1639
Db 468 ProAlaGlyHisAlaValThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAla 487
QY 1640 TTCGTATCAACGCTGAAACAAACACAGAAATCTTCTTCAGGTGATAGAGCAATGTG 1699
Db 488 PheGlyLeuAsnAlaGlnAsnAsnGlnArgIlePheLeuAlaGly---LysLysAsnLeu 506
QY 1700 ATAGACCATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTT 1759
Db 507 ValArgGlnMetAspSerGluAlaLysGluLeuSerPheGlyValProSerLysLeuVal 526
QY 1760 GAGAAGCTCATCAAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTT 1819
Db 527 AspAsnIlePheAsnAsnProAspGluSerTyrPheMetSerPheSerGlnGlnArgGln 546
QY 1820 TCTCAATCTCGTCTCTCTGAGAAAGAGTCTCTCTGAGAAAGAGTCTCTCTGAGAAAGAGGAA 1879
Db 547 -----ArgArgGly-----AsnProLeuAlaSerIleLeu 550

RESULT 7
US-07-955-905A-22
; Sequence 22, Application US/07955905A
; Patent No. 5770433
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: RECOMBINANT 47 AND 31 kD COCOA PROTEINS AND
; NUMBER OF SEQUENCES: 28
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/955,905A
; FILING DATE: 21-JAN-1993
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 566 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Theobroma cacao
; FEATURE:
; NAME/KEY: Protein
```

```
; LOCATION: 1..566
; OTHER INFORMATION: /note= "67 kD Precursor Protein"
US-07-955-905A-22

Alignment Scores:
Pred. No.: 1.75e-69 Length: 566
Score: 811.00 Matches: 195
Percent Similarity: 50.95% Conservative: 127
Best Local Similarity: 30.85% Mismatches: 218
Query Match: 22.77% Indels: 92
Db: 1 Gaps: 19

US-10-728-051-1 (1-2032) x US-07-955-905A-22 (1-566)
QY 65 TCTCCACTGATGCTGCTGCTAGGATCCTTCTCTGCTTCAGTTTCTGCAACCATGCC 124
Db 6 SerProPheIleValLeuIlePheSerLeuLeuLeuSerPheAlaLeuLeuCysSerGly 25
QY 125 AAGTCATCACTTACCAAGAAACAGAGAACCCCTGCGCCAG-----AGTGTC 175
Db 26 ValSerAlaTyrGlyArgLysGlnTyrGluArgAspProArgGlnGlnTyrGluGlnCys 45
QY 176 CTCACAGAGTGTCAACAGGAA---CCGATGACTTGAAGCAAAAGGATGCGAGTCTCGC 232
Db 46 GlnArgArgCysGluSerGluAlaThrGluGluArgGluGlnGlnCysGluGlnArg 65
QY 233 TGCACCAAGCTCGAGTATGATCTCTGTTGTCTATGATCTCTCGAGGACACACTGGCACC 292
Db 66 CysGluArg---GluTyrLysGluGln----- 73
QY 293 ACCAACCAACGTTCCCTCCAGGGAGCGGACACGTCGCCGCCCAACCCGGAGACTACGAT 352
Db 74 -----GlnArgGlnGlnGlnGluGluLeuGlnArgGlnTyrGln----- 86
QY 353 GATGACCGCGCTCAACCCCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 412
Db 87 -----GlnCysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 100
QY 413 GAGCGTGAA-----AGAGAAGAACTGGAGACCAACCAAGAGAGATGGAGG----- 460
Db 101 GluGlnGlnGlnCysGlnArgLysCysTrpGluGlnTyrLysGluGlnGlnGlnGlnGln 120
QY 461 CGACCAAGTCATCAGCAGCAGCAGGAAATAAGCCCGGAGAGAGAGAGAGAGAGAGAGAGAG 520
Db 121 HisGluAsnTyrHisAsnHisLysLysAsnArgSerGluGluGluGluGlnGlnGln--- 139
QY 521 TGGGGAACACAGGTAGCCATGTGAGGAGAGAAACATCTCGGAAACACCTTTCTACTTC 580
Db 140 -----ArgAsnAsnProTyrTyrPhe 146
QY 581 CGCTCAAGCGG---TTTAGCACCCGCTACGGGAACCAACCGGTAGGATCCGGTCTCTG 637
Db 147 ProLysArgArgSerPheGlnThrArgPheArgAspGluGluGluGluGluGluGluGlu 166
QY 638 CAGAGGTTTGACCAAGGTCAAGGCGAGTTTCCAGAAATCTCCAGAAATCACCGTATTGTGAG 697
Db 167 GlnArgPheAlaGluAsnSerProProLeuLysGlyIleAsnAspTyrArgLeuAlaMet 186
QY 698 ATCAGGCGCAACCTAACTCTGTTGTTTCCAGACGCTGATGCTCTATTAACATCCTT 757
Db 187 PheGluAlaAsnProAsnThrPheIleLeuProHisHisCysAspAlaGluAlaIleTyr 206
QY 758 GTTATCCAGCAGGCGCAAGCCCGGTAGCCAAATGGCAATTAACAGAAAGAGCTTT 817
Db 207 PheValThrAsnGlyLysGlyThrIleThrPheValThrHisGluAsnLysGluSerTyr 226
QY 818 AATCTTGACGAGGCGCATGCATCAGATCCCATCGGTTTCATTTCTTACATCTTGAAC 877
Db 227 AsnValGlnArgGlyThrValValSerValProAlaGlySerThrValTyrValValSer 246
QY 878 CGCCATGACAAACAGAACCTCAGAGTAGTAAATCTCCATGCCCTTAAACACACCCGGC 937
Db 247 GlnAspAsnGlnGluLysLeuThrIleAlaValLeuAlaLeuProValAsnSerProGly 266
```

```
QY 938 CAGTTTCAGGATTTCTCCGGGAGCAGCGGAGCAACATCATCTACTTCTGCGGGCTTC 997
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
Db 267 LysTyrGluLeuPheProAlaGlyAsnAsnLysProGluSerTyrGlyAlaPhe 286
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
QY 998 AGCAGGAATACGTTGAGGCGCTTCAATCCGGAATCAATGAGATACGAGGGTCTG 1057
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 287 SerTyrGluValLeuThrValPheAsnThrGlnArgGluLysLeuGluGluLeu 306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1058 TTAGAAGAGAATGACGAGGTGAGCAAGAGAGAGAGAGGCGGAGCGCATGGAGTACTCGG 1117
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
Db 307 -----GluGluGlnArgGlyGlnLysArg----- 314
QY 1118 AGTAGTGAGAACATGAGAGGTGATAGTCAAGTGTCAAGAGACGTTGTAAGACTT 1177
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
Db 315 ---GlnGlnGlyGlnGlnGlyMetPheArgLysAlaLysProGluGlnIleArgAlaIle 333
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
QY 1178 ACTAAGCACCTAAATCCGCTCAAGAAAGAGGTCCGCAAGAGAGGGAGATATACCAAC 1237
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
Db 334 SerGlnGlnAlaThrSerProArgHisArgGlyGlyGluArgLeuAla----- 349
QY 1238 CCAATCAACTTGAGAGAGCGGAGCCGATCTTTCTAACCACTTTGGGAAAGTTATTGAG 1297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 350 ---IleAsnLeuLeuSerGlnSerProValTyrSerAsnGlnAsnGlyArgPhePheGlu 368
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1298 GTGAAGCCAGACAAGAAACCCCGCTTCAGACCTCGACATGATGCTCACTGTGTA 1357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 369 AlaCysProGluAspPheSer--GlnPheGlnAsnMetAspValAlaValSerAlaPhe 387
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1358 GAGATCAAGAGGAGCTTTGATGCTCCACACTTCAACTCAAGGCCATGGTTATCGTC 1417
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
Db 388 LysLeuAsnGlnGlyAlaIlePheValProHisTyrAsnSerLysAlaThrPheValVal 407
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
QY 1418 GTCTCAACAAAGAACTGGAACCTTCAACTCTGCTGTGAAGAAAAGCAACAA-- 1474
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 408 PheValThrAspGlyTyrGlyTyrAlaGlnMetAlaCysProHisLeuSerArgGlnSer 427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1475 -----CAGAGGGAACGGCGGGAAGAAGAGAGCAGCAAGCAAGAGAGGAGGA 1525
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
Db 428 GlnGlySerGlnSerGlyArgGlnAspArgArgGlnGlnGluGluSerGluGluGlu 447
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1526 AGTAACAGAGAGGTGCGTAGGTACACAGCGAGGTGTAAGAGCGGATGTTTCATCATG 1585
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
Db 448 ThrPheGlyGluPheGlnGlnValLysAlaProLeuSerProGlyAspValPheValAla 467
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1586 CCAGCAGCTCATCAGTAGCCATCAACGCTTCTCCGAA-----CTCCATCTGCTTGGC 1639
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 468 ProAlaGlyHisAlaValThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAla 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1640 TTCGGTATCAACGCTGAAAACAACACAGAAATCTCTTCAGGTGATGAAGCAATGTG 1699
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 488 PheGlyLeuAsnAlaGlnAsnAsnGlnArgIlePheLeuAlaGly--LysLysAsnLeu 506
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1700 ATAGACCAGATAGAGAAGCAAGCAAGGATTAGCATTCCTCGGTGCGGTGCAACAGTT 1759
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
Db 507 ValArgGlnMetAspSerGlnAlaLysLeuGluSerPheGlyValProSerLysLeuVal 526
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1760 GAGAAGCTCATCAAAAACCAAGAGAAATCTCACTTTGTGAGTGTCTGCTCTCAATCTCAA 1819
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
Db 527 AspAsnIlePheAsnAsnProAspGluSerTyrPheMetSerPheSerGlnGlnArgGln 546
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1820 TCTCAATCTCGTCTCTCGTGAAGAGTCTCTCTGAGAAAGAGTCTCTGAGAAAGAGGAGAA 1879
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
Db 547 -----ArgArgAspGlu 550
QY 1880 AACCAAGGAGGAGGTCCACTCTCTTCAATTTTG 1915
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
Db 551 ArgArgGly-----AsnProLeuAlaSerIleLeu 560
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
```

RESULT 8

US-07-955-905A-27

; Sequence 27, Application US/07955905A

; Patent No. 5770433

```
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: RECOMBINANT 47 AND 31 kD COCOA PROTEINS AND
; TITLE OF INVENTION: PRECURSOR
; NUMBER OF SEQUENCES: 28
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/955,905A
; FILING DATE: 21-JAN-1993
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 421 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Phaseolus vulgaris
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..421
; OTHER INFORMATION: /note= "Vicilin from P. vulgaris"
; US-07-955-905A-27
```

```
Alignment Scores:
Pred. No.: 2,46e-67 Length: 421
Score: 789.00 Matches: 182
Percent Similarity: 58.44% Conservative: 88
Best Local Similarity: 39.39% Mismatches: 106
Query Match: 22.12% Indels: 86
DB: 1 Gaps: 14
```

US-10-728-051-1 (1-2032) x US-07-955-905A-27 (1-421)

```
QY 548 GAAGAAACATCTCGGAACAAACCTTCTTACTTCCCGTCA---AGCGGTTTATGACCCCGC 604
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
Db 30 GluGluGluSerGlnAspAsnProPheTyrPheAsnSerAspAsnSerTrpAsnThrLeu 49
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
QY 605 TACGGGAACCAAAACCGCTAGGTAGTCCGGGTCTCTCAGAGGTTTACCAAGGTCAAGCAG 664
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
Db 50 PheLysAsnGlnTyrGlyHisIleArgValLeuGlnArgPheAspGlnGlnSerLysArg 69
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
QY 665 TTTTCAGATCTCCAGATCACCGTATTGTGCAGATCGAGCCAAACCTTAACACTCTTTGTT 724
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 70 LeuGlnAsnLeuGluAspTyrArgLeuValGluPheArgSerLysProGluThrLeuLeu 89
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 725 CTTCCCAAGCAGCGCTGATGCTGATAACATCTTGTATTATCCAGCAAGGGCAAGCCCGTG 784
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 90 LeuProGlnGlnAlaAspAlaGluLeuLeuLeuValValArgSerGlySerAlaIleLeu 109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 785 ACCGTAGCAATGGCAATACAGAAAGAGC---TTTAATCTTTCAGAGGGCCATGCACCTC 841
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
Db 110 ValLeuValLysProAspArgArgGluTyrPheLeuThrSerAspAsnProIle 129
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
QY 842 -----AGAATCCCATCCGGTTTTCATTCTTCTTACATCTTGAACCGCCATGAC 886
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
Db 130 PheSerAspHisGlnLysIleProAlaGlyThrIlePheTyrLeuValAsnProAspPro 149
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
QY 887 AACCAAGACCTCAGAGTAGCTAAATCTCCATCCCGCTTAACACACCCCGCCAGTTTGAG 946
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
Db 150 LysGluAspLeuArgIleIleGlnLeuAlaMetProValAsnAsnPro---GlnIleHis 168
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
QY 947 GATTCTTCCCGGCGAGCGGAGCACCATCATCTTCTTCTGAGGGCTTCAGCAGGAAT 1006
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
Db 169 GluPhePheLeuSerThrGluAlaGlnGlnSerTyrLeuGlnGluPheSerLysHis 188
    : : : : : ||| ||| ||| ||| ||| : : : ||| ||| ||| ||| |||
QY 1007 ACCTTGGAGCGCCCTTCAATCGGAATTCAAATGAGATCGGAGGTGCTGTAGAGAG 1066
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 189 IleLeuGluAlaSerPheAsnSerLysPheGluGluIleAsnArgValLeuPheGluGlu 208
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



```
QY 752 ATCTTTGTTATCCAGCAAGGCAAGCCAGCTGACCGTAGCAATATGCAATATACAGAAAG 811
Db 241 ILeTyrLeuValThrAsnGlyArgGlyThrLeuThrPheLeuThrHisGluAsnLysGlu 260
QY 812 AGCTTTAATCTTACGAGGCGCATCCAGTCCAGTCCCGTTTCATTTCTCTACATC 871
Db 261 SerTyrAsnValValProGlyValValValArgValProAlaGlySerThrValTyrLeu 280
QY 872 TTGAACCCCATGACACCAACCAACCTCAGATAGTAAATCTCCATGCCCGTTTAAACACA 931
Db 281 AlaAsnGlnAspAsnLysGluLysLeuLeileleAlaValLeuHisArgProValAsnAsn 300
QY 932 CCCGCCAGTTTGGAGTATTTCTCCCGCGAGAGCCGAGACCAATCATCTCTACTTGCAG 991
Db 301 ProArgGlnPheGluGluPhePheProAlaGlySerGlnArgProGlnSerTyrLeuArg 320
QY 992 GGCTTTCAGCAGGAATAGTTGGAGCGCTTCAATCGGAATTCATGAGATACGGAGG 1051
Db 321 AlaPheSerArgGluLeuLeuGluProAlaPheAsnThr-----ArgSer 335
QY 1052 GTGCTGTTAGAAAGAAATGAGAGGTGAGCAAGAGAGAGAGCGGAGCGATGGAGT 1111
Db 336 GluGlnLeuAspGluLeuPheGlyArgGlnSerHisArgArgGlnGln----- 352
QY 1112 ACTCGGAGTAGTGAGAACATGAGAGTAGTATAGTCAAAGTGTCAAAGAGACGTTGAA 1171
Db 353 -----GlyGlnGlyMetPheArgLysAlaSerGlnGluGlnIleArg 366
QY 1172 GAACCTTACTAAGCAGCTAAATCGTCTCAAGAAAGCTCCGGAAGAGAGGGAGATATC 1231
Db 367 AlaLeuSerGlnGluAlaThrSerProArgGluLysSerGlyGluArgPheAla----- 384
QY 1232 ACCAACCCCAATCACTTTGAGAGAGCGGAGCGGATCTTTCTAACAACTTTGGGAAGTTA 1291
Db 385 -----PheAsnLeuLeuTyrArgThrProArgTyrSerAsnGlnAsnGlyArgPhe 401
QY 1292 TTTGAGTGAAGCCAGACAGAAAGACCCCGAGCTTCAGACCTGGACATGATGCTCAAC 1351
Db 402 TyrGluAlaCysPro---ArgGluPheArgGlnLeuSerAspIleAsnValThrValSer 420
QY 1352 TGTGTAGAGATCAAAAGAGAGCTTGTATGCTCCACACTTCAACTCAAAAGGCCATGGTT 1411
Db 421 AlaLeuGlnLeuAsnGlnGlySerIlePheValProHisTyrAsnSerLysAlaThrPhe 440
QY 1412 ATCGTCTGCTCAACAAAGAACTGGAAACCTTGAACCTCGGTGTGAAGAAAGACAA 1471
Db 441 ValValLeuValAsnGluGlyAsnGlyTyrValGluMetValSerProHisLeuProArg 460
QY 1472 CAACAGAGGGAGCGCGGGAAGAGAGGAGGACGAGACGAAAGAGAGAGGAGGAAGTAAC 1531
Db 461 GlnSerSerPheGluGluGluGluGluGlnGlnGlnGlnGlnGlnGluGluArg 480
QY 1532 AGA-----GAGGTGCTAGGTATACAGACGAGGTGAAGAGCGGATGTTCATCATG 1585
Db 481 ArgSerGlyGlnTyrArgLysIleArgSerGlnLeuSerArgGlyAspIlePheValVal 500
QY 1586 CCAGCAGCTCATCCAGTAGCCATCAACGCTTCC-----TCCGAACCTCCATCTGCTTGGC 1639
Db 501 ProAlaAsnPheProValThrPheValAlaSerGlnAsnGlnAsnLeuArgMetThrGly 520
QY 1640 TTCGGT-----ATCAACGCTGAAACCAACACAGAAATCTTCTTCTGTCAGGT 1684
Db 521 PheGlyLeuTyrAsnGlnAsnIleAsnProAspHisAsnGlnArgIlePheValAlaGly 540
QY 1685 GATAGACATATGTATAGACCATAGACATAGACGAGGAGGATTTAGCATTCCTCCGTTGG 1744
Db 541 ---LysIleAsnHisValArgGlnTrpAspSerGlnAlaLysGluLeuAlaPheGlyVal 559
QY 1745 TCGGGTGAACAAGTTGAGAGAGCTCATCAAAACCCAG--AAGGAATCTCACTTTGTGAGT 1801
Db 560 SerSerArgLeuValAspGluIlePheAsnAsnAsnProGlnGlnSerTyrPheValSer 579
```

```
QY 1802 GCTCGTCCTCAATCTCAATCTCAA 1825
Db 580 ArgGlnArgGlnArgAlaSerGlu 587
```

RESULT 10

```
US-09-424-283-3
; Sequence 3, Application US/09424283
; Patent No. 6437219
; GENERAL INFORMATION:
; APPLICANT: Grimes, et al.
; TITLE OF INVENTION: Sucrose binding proteins
; FILE REFERENCE: 4630-50206
; CURRENT APPLICATION NUMBER: US/09/424,283
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: PCT/US98/10465
; PRIOR FILING DATE: 1998-05-21
; PRIOR APPLICATION NUMBER: US 60/047,568
; PRIOR FILING DATE: 1997-05-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 489
; TYPE: PRT
; ORGANISM: Glycine max
US-09-424-283-3
```

```
Alignment Scores:
Pred. No.: 2,99e-48 Length: 489
Score: 591.50 Matches: 156
Percent Similarity: 43.99% Conservative: 100
Best Local Similarity: 26.80% Mismatches: 195
Query Match: 16.61% Indels: 131
DB: 4 Gaps: 13
```

US-10-728-051-1 (1-2032) x US-09-424-283-3 (1-489)

```
QY 71 CTGATCTGTTGCTAGGATCTTGTCTCGCTTCTGCTTCTCAACGATGCAAGTCA 130
Db 16 LeuLeuAlaLeuIleSerAsnLeuAlaLeuGlyLysLeuLysGluThrGluValGlu--- 34
QY 131 TCACCTTACCAGAAGAAAACAGAAACCCCTGCGCCAGAGGTGCTTCCAGAGTTGTCAA 190
Db 35 -----GluAspProGluLeuValThrCysLysHisGlnCysGln 47
QY 191 CAGGAACCGGAT-----GACTTGAAGCAAAAGCATGCGAGTCTCGCTGCACCAAGCTC 244
Db 48 GlnGlnArgGlnTyrThrGluSerAspLysArgThrCysLeuGlnGlnCysAspSerMet 67
QY 245 GAGTATGATCTCTGTTGTGTCTATGATCTCTGAGGACACACTGGCACCACTGGCAACG 304
Db 68 LysGlnGlu----- 70
QY 305 TCCCTCTCCAGGGGAGCGGACACGTTGCGCCCAACCCGGAGACTACGATGATGATGATG 364
Db 70 ----- 70
QY 365 CAACCCCGAAGAGAGGAGGAGCGCCGATGGGACCACTGGAGCGGAGGCGGTGAAGA 424
Db 71 -----ArgGluLysGlnVal 75
QY 425 GAAGAAGACTGGAGACAACAAGAGAAGATTTGAGGCGGACCAAGTATCATGACGACCGG 484
Db 76 GluGluGluThrArgGluLysGluGluGlu-----HisGlnGlnGln--- 89
QY 485 AAATAAGCCCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 544
Db 89 ----- 89
QY 545 AGGGAAGAAACATCTCGGAACCAACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 601
Db 90 HisGluGluGluGluAspGluAsnProTyrValPheGluGluAspLysAspPheSerThr 109
QY 602 CGCTACGGGAACCAAAACCGTAGGATCCGGGTCTCTGAGAGGTTTGTACCAAGGTCAAG 661
```

Db 110 ArgValGluThrGluGlyGlySerIleArgValLeuLysLysPheThrGluLysSerLys 129
QY 662 CAGTTTCAGAAATCCAGAAATCAGCGTATTGTGCAGATCGAGGCCAAACCTACACTCTT 721
Db 130 LeuLeuGlnGlyIleGluAsnPheArgLeuAlaIleLeuGluAlaArgAlaHisThrPhe 149
QY 722 GTTCTTCCCAAGCAGCTGATGCTGATAACATCTTTGTTATCCAGCAAGGGCAAGCCACC 781
Db 150 ValSerProArgHisPheAspSerGluValValLeuPheAsnIleLysGlyArgAlaVal 169
QY 782 GTACCGTAGCAATGGAATATACAGAAAGAGCTTTAATCTTACAGAGGCCCATGCACTC 841
Db 170 LeuGlyLeuValArgGluSerGluThrGluLysIleThrLeuGluProGlyAspMetIle 189
QY 842 AGAATCCATCCCGTTTCATCTTCTCATCTTCAACCGCATCATCACCAACCAACCTCAGA 901
Db 190 HisIleProAlaGlyThrProLeuTyrlleValAsnArgAspGluAsnGluLysLeu 209
QY 902 GTAGCTAAATCTCCATGCCCGTTAAACACACCCGGCAGTTTGAGGATTTCTTCCGGCG 961
Db 210 LeuAlaMetLeuHisIleProValSerThrProGlyLysPheGluGluPheGlyPro 229
QY 962 AGCAGCGGACCAATCATCTACTTGCAGGGCTTTCAGCAGGAATAGCTTGGAGCGGCC 1021
Db 230 GlyGlyArgAspProGluSerValLeuSerAlaPheSerTrpAsnValLeuGlnAla 249
QY 1022 TTCAATCGCGAATTCATAGATACGAGGGTCTGTAGAGAGAGATGCAAGAGTGAG 1081
Db 250 LeuGlnThrProLysGlyLysLeuGluArgLeuPhe----- 261
QY 1082 CAAGAGGAGAGGCGAGAGCGATGAGTACTCGAGTAGTGAGAACCAATGAAGAGTG 1141
Db 262 -----AsnGlnGlnAsnGluGlySer 268
QY 1142 ATAGTCAAGTGTCAAGAGGACGTTGAGAACTTACTAAGCAGCGTAATCCGTCTCA 1201
Db 269 IlePheLysIleSerArgGluArgValArgAlaLeu-----AlaProThr 283
QY 1202 AAGAAGGCTCC-----GAAGAAGAGGAGATATCACCAACCAATCACTTGAGAGAA 1255
Db 284 LysLysSerSerTrpProPheGlyGlySerLysAlaGlnPheAsnIlePheSer 303
QY 1256 GCGAGCGCGATCTTCTAACTTTGGAAGTTATTGAGTGAGCCAGACAGAAAG 1315
Db 304 LysArgProThrPheSerAsnGlyTyrglyArgLeuThrGluValGlyProAspAspGlu 323
QY 1316 AACCCCGAGCTCAGGACCTGACATGATCTCACCTGTGTAGATCAAGAGGAGCT 1375
Db 324 LysSerTrpLeuGlnArgLeuAsnLeuMetLeuThrPheThrAsnIleThrGlnArgSer 343
QY 1376 TTGATGCTCCCACTTCAACTCAAGGCCATGTTATCGCTCGTCAACAAAGGAAC 1435
Db 344 MetSerThrIleHisTyrlAsnSerHisAlaThrLysIleAlaLeuValMetAspGlyArg 363
QY 1436 GGAACCTTGAACCTCGGTGCTGTGAAGAAAAGACAAACAGAGGGGACGCGGGAAGAA 1495
Db 364 GlyHisLeuGlnIleSerCysProHisMetSerSerArg----- 376
QY 1496 GAGGAGACCAAGCAAGAGAGGAGGAGGAGGAGTAAACAGAGGTCGTAGGTACACAGCG 1555
Db 377 -----SerAspSerLysHisAspLysSerSerProSerTyrlHisArgIleSerAla 393
QY 1556 AGGTTGAAGAGCGGATGTTTCATCATCGCCAGCCTCATCTCAGTAGCCATCAACGCT 1615
Db 394 AspLeuLysProGlyMetValPheValProProGlyHisProPheValThrIleAla 413
QY 1616 TCCTCCGAATCCATCTGCTT-----GGCTTGGTATCAACGCTGAAAACACACAGAGA 1669
Db 414 SerAsnLysGluAsnLeuLeuIleCysPheGluValAsnValArgAspAsnLysLys 433
QY 1670 ATCTTCTTCAGTGATAGGACAATGTATAGACCAGATAGAGAGCAAGCGAGGAT 1729

Db 434 PheThrPheAlaGly---LysAspAsnIleValSerSerLeuAspAsnValAlaLysGlu 452
QY 1730 TTAGCATTCCTCGGTGGGTGAGCAACAAAGTTGAGAAGCTCATCAAAAACCCAGAGGAATCT 1789
Db 453 LeuAlaPheAsnTyrlProSerGluMetValAsnGly---ValSerGluArgLysGluSer 471
QY 1790 CACTTT 1795
Db 472 LeuPhe 473
RESULT 11
US-09-424-283-1
; Sequence 1, Application US/09424283
; Patent No. 6437219
; GENERAL INFORMATION:
; APPLICANT: Grimes, et al.
; TITLE OF INVENTION: Sucrose binding proteins
; FILE REFERENCE: 4630-50206
; CURRENT APPLICATION NUMBER: US/09/424,283
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: PCT/US98/10465
; PRIOR FILING DATE: 1998-05-21
; PRIOR APPLICATION NUMBER: US 60/047,568
; PRIOR FILING DATE: 1997-05-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 524
; TYPE: PRT
; ORGANISM: Glycine max
US-09-424-283-1
Alignment Scores:
Pred. No.: 4,73e-46 Length: 524
Score: 569.00 Matches: 163
Percent Similarity: 44.03% Conservative: 106
Best Local Similarity: 26.68% Mismatches: 208
Query Match: 15.97% Indels: 134
Gaps: 17
US-10-728-051-1 (1-2032) x US-09-424-283-1 (1-524)
QY 50 ATGAGAGGAGGAGTTCTCCACTGATGCTGTGTAGGATCCTTGTCTGCTTCAGTT 109
Db 3 MetArgThrLysLeuSerLeuAlaIlePhePhePheLeuLeuAlaLeuPheSer--- 21
QY 110 TCTGCAACGATGCCAAGTCATCACCTTACCAGAGAAACAGAGAACCCCTGCGCCAG 169
Db 22 AsnLeuAlaPheGlyLysCysLysGluThrGluValGluGluAspProGluLeuVal 41
QY 170 AGGTGCTCCAGAGTTGTCAACAGAGAACCGGAT-----GACTTGAAGCAAGGATGC 223
Db 42 ThrCysLysHisGlnCysGlnGlnGlnGlnGlnTyrlThrGluGlyAspLysArgValCys 61
QY 224 GAGTCTCGCTGCACCAAGCTCAGTATGATCTCGTTGTGTCTATCATCTCGAGGACAC 283
Db 62 LeuGlnSerCys----- 65
QY 284 ACTGGCACCAACCAACCAAGTTTCCCTCCAGGGAGCGGACACGTCGCGCCCAACCCGGA 343
Db 65 ----- 65
QY 344 GACTACGATGATGACGCGCTCAACCCGAGAGAGAGAGAGGAGCGGATGGGACCGAGT 403
Db 66 -----AspArgTyrlHisArgMetLysGlnGlu----- 74
QY 404 GGACCGAGGAGCGGTCAAGAGAGAGAGAGACTGGAGACCAACCAAGAGAGATGGAGCGCA 463
Db 75 -----ArgGluLysGlnIleGlnGluGluThrArgGluLysGluGlu----- 89
QY 464 CCAAGTCATCAGCAGCCACCGAAATAAGGCCCGGAGAGAGAGAGAGAGCAAGATGG 523
Db 90 -----GluSerArgGluArgGluGlu----- 97

Qy	1586	CCAGCAGCTCATCCAGTAGCCATCAACGCTTCTCCGAACTCCATCTGCTT-----GGC	163
Db	419	ProProGlyHisProPheValThrIleAlaSerAsnLysGluAsnLeuLeuMetIleCys	438
Qy	1640	TTCGGTATCAAGCGTGAACCAACACAGAAATCTTCCTTCGAGGTGATAAGGACAAATGTG	1699
Db	439	PheGluValAsnAlaAargAspAsnLysPheThrPheIaGly---LysAspAsnIle	457
Qy	1700	ATAGACCAGATGAGAAAGCAAGCGAAGATTTTAGCATTTCCCTCGGTCGGGTGAACAAAGTT	1755
Db	458	ValSerSerLeuAspAsnValAlaLysGluLeuAlaPheAsnTyrProSerGluMetVal	477
Qy	1760	-----GAGAACTCATC-----	1771
Db	478	AsnGlyValPheLeuLeuGlnAargPheLeuGluAargLysLeuIleGlyAargLeuTyrHis	497
Qy	1772	-----AAAAACACAGAAAGGAATCTCACTTT	1795
Db	498	LeuProHisLysAspAargLysGluSerPhePhe	508
RESULT 12			
US-09-323-195A-18			
; Sequence 18, Application US/09323195A			
; Patent No. 6462257			
; GENERAL INFORMATION:			
; APPLICANT: Pullman, Gerald			
; APPLICANT: Cairney, John			
; APPLICANT: Perrera, Ranjan			
; TITLE OF INVENTION: VICILIN-LIKE SEED STORAGE PROTEIN GENE PROMOTER AND			
; TITLE OF INVENTION: METHODS OF USING THE SAME			
; FILE REFERENCE: IPST0009			
; CURRENT APPLICATION NUMBER: US/09/323,195A			
; CURRENT FILING DATE: 1999-06-01			
; NUMBER OF SEQ ID NOS: 19			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 18			
; LENGTH: 448			
; TYPE: PRT			
; ORGANISM: Picea glauca			
US-09-323-195A-18			
Alignment Scores:			
Pred. NO.: 3.57e-45 Length: 448			
Score: 559.50 Matches: 134			
Percent Similarity: 50.12% Conservative: 82			
Best Local Similarity: 31.09% Mismatches: 178			
Query Match: 15.71% Indels: 37			
DB: 4 Gaps: 10			
US-10-728-051-1 (1-2032) x US-09-323-195A-18 (1-448)			
Qy	545	AGGGAAGAAACATCTCGGAACACCCITTTCTACTTCCGTCAGGCGGTTTAGCACCCGC	604
Db	44	ArgGluGluGluAargGluGluAsnProTyrValPheHisSerAppSerPheAargThrAarg	63
Qy	605	TACGGGAACCAAAACGGGTAGGATCCGGGTCTCGAGAGGTTTGACCAAGGTCAGGCGAG	664
Db	64	AlaSerSerGluAlaGlyLysIleAargAlaLeuProAsnPheGlyGluValSerGluLeu	83
Qy	665	TTTCCGAATCTCCAGAATCACCGGTATTGTGCAGATTCGAGGCCAACCTTAACACTCTTGT	724
Db	84	LeuGluGlyIleAargLysPheAargValThrCysIleGluMetLysProAsnThrValMet	103
Qy	725	CTTCCCAAGCAGCGTAGCTGCTGATAACATCTTGTATCCAGCAAGGCGCAAGCCACCGTG	784
Db	104	LeuProHisTyrIleAspAlaThrTrpIleLeuTyrValThrAargGlyAargGlyTyrIle	123
Qy	785	ACCGTAGCAATGGCAATAACAGAAAGAGCTTTAATCTTGACGAGGCGCATGCACCTCAGA	844
Db	124	AlaTyrValHisGlnAsnGluLeuValLysAargLysLeuGluGlyAspValPheGly	143
Qy	845	ATCCCATCCGGTTTCATTTCTTCATCTTGAAACCGCCATGACACACAGAACTCAGAGTA	904

```
Db 144 ValProSerGlyHisThrPheTyrLeuValAsnAspHisAsnThrLeuArgIle 163
QY 905 GCTAAATCTCCATGCCCCGTTAAACACA---CCGGCCAGTTTCTGAGGATTTCTCCGGCG 961
Db 164 AlaSerLeuValArgProValSerThrValArgGlyGluTyrGlnProPheTyrValAla 193
QY 962 AGACGCCGAGACCATCATCTCTGACGGCTTACAGAGGATACGTTGGAGCGGCC 1021
Db 184 GlyGlyArgAsnProGlnThrValTyrSerAlaPheSerAspValLeuGluAlaA 203
QY 1022 TTCATGTCGGAATTCATAGATACGAGGGTCTGTGTAGAGATGACGAGGTGAG 1081
Db 204 PheAsnThrAsnValGlnGluLeuGluArgIlePhe-----GlyGly--- 217
QY 1082 CAAGAGGAGAGGCGCAGAGCGATGAGTAGTCTCGGAGTAGTGAGAACATGAAGAGTG 1141
Db 218 -----HisLysSerGlyVal 222
QY 1142 ATAGTCAAGTGTCAAAGGAGCAGCTGTGAAGACTTACTAAGCACGCTAAATCCGTCTCA 1201
Db 223 IleIleHisAlaAsnGluGluGlnIleArgGluMetMetArg-----LysArgGlyPhe 240
QY 1202 AAGAAAGGCTCCGAGAGGAGGAGATATACCAACCCATCACTGAGAGAGGCGAG 1261
Db 241 SerAlaGlySerMetSerAlaProGluHisProLysProPheAsnLeuArgAsnGlnLys 260
QY 1262 CCCGATCTTTTAACTTTGGGAGTTATTTGAGGTGAAGCCAGACAGAAAGAACCC 1321
Db 261 ProAspPheLysAsnGluAsnGlyArg--PheThrIleAlaGlyProLysAsnTyrPro 279
QY 1322 CAGCTTCAGACCTGACATGATGCTCACTGCTGTAGAGATCAAGAGAGGAGCTTTGATG 1381
Db 280 PheLeuAspAlaLeuAspValSerValGlyLeuAlaAspLeuAsnProGlySerMetThr 299
QY 1382 CTCCCACACTTCAACTCAAGGCCATGCTTATCGTCTGCTCAACAAAGAACTGCAAC 1441
Db 300 AlaProSerLeuAsnSerLysThrSerIleGlyIleValThrAsnGlyGluGlyArg 319
QY 1442 CTTGAACCTCGTGTCTGAAGAAAGAGCAACAGAGGGGACGGCGGAAGAGAGAG 1501
Db 320 IleGluMetAlaCysProHisLeuGlyGlnHisGlyTrpSerSerProArgGluArgGly 339
QY 1502 GACGAGACGAAAGAGGAGGAGAAAGTAACAGAGAGTGCTGATAGGTACACAGCGAGTTG 1561
Db 340 AspGlnAsp-----IleThrTyrGlnArgValTrpAlaLysLeu 352
QY 1562 AAGGAAGGCGATGTTTCATCATGCCAGAGCTCATCCAGTAGCCATCAACGTTCC--- 1618
Db 353 ArgThrGlySerValTyrIleValProAlaGlyHisProIleThrGluIleAlaSerThr 372
QY 1619 ---TCCGAATCCATCTGCTTGGCTTCGGTATCAAGCTCAACCAACACAGAACTTTC 1675
Db 373 AsnSerArgLeuGlnIleLeuTrpPheAspLeuAsnThrArgGlyAsnGluArgGlnPhe 392
QY 1676 CTTGCGAGGTATAGGACATGTGATAGCAGATAGAGCAAGCAAGGAGGATTTAGCA 1735
Db 393 LeuAlaGly---LysAsnAsnValLeuAsnThrLeuGluArgGluIleArgGlnLeuSer 411
QY 1736 TTC---CCTGGTCCGGTGAACAAGTTGAGAAGCTCATCAAAACACAGAGGATCTCAC 1792
Db 412 PheAsnValProArgGlyGluGluIleGluValLeuGlnAlaGlnLysAsp---Gln 430
QY 1793 TTTGTAGTGCTCGTCTCAATCTCAATCTCA 1825
Db 431 ValIleLeuArgGlyProGlnArgArgSerArg 441
```

RESULT 13

US-09-424-283-4

; Sequence 4, Application US/09424283

; Patent No. 6437219

; GENERAL INFORMATION:

; APPLICANT: Grimes, et al.

; TITLE OF INVENTION: Sucrose binding proteins

```
; FILE REFERENCE: 4630-50206
; CURRENT APPLICATION NUMBER: US/09/424,283
; PRIOR FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: PCT/US98/10465
; PRIOR FILING DATE: 1998-05-21
; PRIOR APPLICATION NUMBER: US 60/047,568
; PRIOR FILING DATE: 1997-05-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Glycine max
US-09-424-283-4
```

Alignment Scores:

Pred. No.:	7,75e-41	Length:	409
Score:	514.50	Matches:	133
Percent Similarity:	42.72%	Conservative:	87
Best Local Similarity:	25.83%	Mismatches:	168
Query Match:	14.44%	Indels:	127
DB:	4	Gaps:	10

US-10-728-051-1 (1-2032) x US-09-424-283-4 (1-409)

```
QY 71 CTGATGCTCTGTAGGGATCCTGTCTGGCTTCTGCTCAACGATGCCAAGTCA 130
Db 16 LeuLeuAlaLeuIleSerAsnLeuAlaLeuGlyLysLeuLysGluThrGluValGlu--- 34
QY 131 TCACCTTACCAGAGAAACAGAGAACCCCTCGGCCAGAGGTGCTCCAGAGTTGTCAA 190
Db 35 -----GluAspProGluLeuValThrCysLysHisGlnCysGln 47
QY 191 CAGGAACCGGAT-----GACTTGAAGCAAAAGGATCGAGTCTCGCTGCACCAAGCTC 244
Db 48 GlnGlnArgGlnTyrThrGlnSerAspLysArgThrCysLeuGlnGlnCysAspSerMet 67
QY 245 GAGTATGATCTCTGTTGTCTATGATCCTCGAGGACACACTGGCACCACCAACCAACGT 304
Db 68 LysGlnGlu----- 70
QY 305 TCCCTTCCAGGGAGCGGACACGTGGCCGCCAACCCGGAGACTAGATGATGACCGCGT 364
Db 70 ----- 70
QY 365 CAACCCGAGAGAGGAGGAGCGGATGGGACCGAGCTGGACCGAGGAGCGTGAAGA 424
Db 71 -----ArgGluLysGlnVal 75
QY 425 GAAGAAGACTGGAGACAACCAAGAGAAGATTGGAGGCGACCAAGTCAATCAGCAGCACCG 484
Db 76 GluGluGluThrArgGluLysGluGluGlu-----HisGlnGln--- 89
QY 485 AAATAAGCCCGAAGGAGAGAGAGAGAAACAAGAGTGGGGAACACCAAGGTAGCCATGTG 544
Db 89 ----- 89
QY 545 AGGGAAGAAACATCTCGGAACAACCTTCTTACTTC---CCGTCAAGGCGGTTTACAC 601
Db 90 HisGluGluGluGluAspGluAsnProTyrValPheGluGluAspLysAspPheSerThr 109
QY 602 CCCTACGGGAACCAAAACGGTAGGATCCGGGTCTCTGCAGAGGTTTTCACCAAGGTCAAGG 661
Db 110 ArgValGluThrGluGlyGlySerIleArgValLeuLysLysPheThrGluLysSerLys 129
QY 662 CAGTTTTCAGATCTCCAGAAATCACCATTGTCAGATCGAGGCCAAACCTTAACACTCTT 721
Db 130 LeuLeuGlnGlyIleGluAsnPheArgLeuAlaIleLeuGluAlaArgAlaHisThrPhe 149
QY 722 GTTCTTCCCAAGCAGCTGATGCTATAACATCTTGTATCCAGCAAGGGCAAGCCACC 781
Db 150 ValSerProArgHisPheAspSerGluValValLeuPheAsnIleLysGlyArgAlaVal 169
```

```
QY 782 GTGACCGTAGCAATGCAATAACAGAAAGAGCTTTAACTTGTACGAGGGCCATGCACCTC 841
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
170 LeuGlyLeuValArgGluSerGluThrGluLysIleThrLeuGluProGlyAspMetIle 189
QY 842 AGAATCCCATCCGGTTTCATTTCTACATCTTGAACCGCCATGACACACCAACCTCAGA 901
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
190 HisIleProAlaGlyThrProLeuTyrlleValAsnArgAspGluAsnGluLysLeuLeu 209
QY 902 GTAGCTAAATCTCCATCCCGCTTAAACACACCGCCGAGTTTGAGGATTTCTTCCCGGCG 961
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
210 LeuAlaMetLeuHisIleProValSerThrProGlyLysPheGluGluPhePheGlyPro 229
QY 962 AGCAGCCGAGACCAATCATCTCTACTTCAGGGCTTCAGCAGGAATACGTTGGAGCGGCC 1021
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
230 GlyGlyArgAspProGluSerValLeuSerAlaPheSerTrpAsnValLeuGlnAlaAla 249
QY 1022 TTCATCCGAATTCATAGATACGAGGAGGTGCTGTAGAGAAGAAATGCAGGAGGTGAG 1081
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
250 LeuGlnThrProLysGlyLysLeuGluArgLeuPhe----- 261
QY 1082 CAAGAGGAGAGCGCAGAGCGGAGTACTCGGAGTAGTGAGAACAAATGAAGGAGTG 1141
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
262 -----AsnGlnGlnAsnGluGlySer 268
QY 1142 ATAGTCAAGTGTCAAGGAGCAGCTTGAAGAACTTACTAAGCAGCTAAATCGTCTCA 1201
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
269 IlePheLysIleSerArgGluArgValArgAlaLeu-----AlaProThr 283
QY 1202 AAGAAGGCTCC-----GAAGAAGAGGAGATATACCAACCAATCAACTTGAGAGAA 1255
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
284 LysLysSerSerTrpTrpProPheGlyGlySerLysAlaGlnPheAsnIlePheSer 303
QY 1256 GCGAGCCGATCTTTCTAACTTTGGGAAGTTATTTGAGGTGAAGCCAGACAAAGAAG 1315
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
304 LysArgProThrPheSerAsnGlyTyrglyArgLeuThrGluValGlyProAspAspGlu 323
QY 1316 AACCCCGAGCTTACGACCTGACATGATCTACCTGTTGAGATCAAGAAGAGACT 1375
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
324 LysSerTrpLeuGlnArgLeuAsnLeuMetLeuThrPheThrAsnIleThrGlnArgSer 343
QY 1376 TTGATGCTCCACACTTCAACTCAAGGCCATGTTATCTGTCGTCGTCACAAAGGAAGT 1435
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
344 MetSerThrIleHisTyrsAsnSerHisAlaThrLysIleAlaLeuValMetAspGlyArg 363
QY 1436 GGAACCTTGAACCTCGTGGCTGTGAAGAAAGAGCAACACAGAGGGGACGCGGGAAGAA 1495
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
364 GlyHisLeuGlnIleSerCysProHisMetSerSerArg----- 376
QY 1496 GAGGAGCAAGCAAGCAAGAGGAGGAGGAAGTAAACAGAGAGGTGCGTAGGTACACAGCG 1555
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
377 -----SerAspSerLysHisAspLysSerSerProSerTyHisArgIleSerAla 393
QY 1556 AGGTTGAAGGAGCGATGTTTCATCATCCATCCAGCAGCTCATCCA 1600
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
394 AspLeuLysProGlyMetValPheValProProGlyHisPro 408
```

RESULT 14

```
US-09-323-195A-17
; Sequence 17, Application US/09323195A
; Patent No. 6462257
; GENERAL INFORMATION:
; APPLICANT: Pullman, Gerald
; APPLICANT: Cairney, John
; APPLICANT: Perrera, Ranjan
; TITLE OF INVENTION: VICILIN-LIKE SEED STORAGE PROTEIN GENE PROMOTER AND
; TITLE OF INVENTION: METHODS OF USING THE SAME
; FILE REFERENCE: IPST0009
; CURRENT APPLICATION NUMBER: US/09/323.195A
; CURRENT FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 523
```

```
; TYPE: PRT
; ORGANISM: Pinus taeda
US-09-323-195A-17

Alignment Scores:
Pred. No.: 1,13e-40 Length: 523
Score: 513.50 Matches: 151
Percent Similarity: 46.23% Conservative: 88
Best Local Similarity: 29.21% Mismatches: 215
Query Match: 14.42% Indels: 63
Db: 4 Gaps: 16

US-10-728-051-1 (1-2032) x US-09-323-195A-17 (1-523)
QY 533 GGTAGCCATGTGTGAGGAAGAAACATCTCGAAACAACCTTTCTACTTCCGTCGAAGCGG 592
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
43 GlyHisGlnArgArgGluGluArgGluGluAsnProTyValPheHisSerAspArg 62
QY 593 TTTAGCACCGCTACGGGAACCAACGAGTAGAGTACGGGTCTCGCAGAGGTTTGACCAA 652
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
63 PheArgMetArgAlaSerSerAspAlaGlyGluIleArgAlaLeuProAsnPheGlyGlu 82
QY 653 AGGTCAGGCGAGTTTCCAGAAATCTCCAGAATCCCGTATTGTGCAGATCCGAGGCCAAACCT 712
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
83 AlaSerGluLeuLeuGluGlyIleSerLysTyArgValThrCysIleGluMetArgPro 102
QY 713 AACACTCTTGTCTTCCCAAGCACGCTGATGCTGATAACATCTTGTATCCAGCAAGG 772
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
103 AsnThrValMetLeuProHisTyLeuAspAlaThrTrpIleLeuTyValThrGlyGly 122
QY 773 CAAGCCACGCTGACCGTAGCAATGCAATACAGAAAGAGCTTTAATCTTGACGAGGGC 832
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
123 ArgGlyTyrlleAlaTyValHisGlnAsnGluLeuValLysArgLysLeuGluGluGly 142
QY 833 CATGCACTCAGAAATCCATCCCGGTTTCATTTCTCTACATCTTGAACCGCCATCAACAG 892
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
143 AspValPheGlyValProSerGlyHisThrPheTyLeuValAsnAsnAspAspHisAsn 162
QY 893 AACCTCAGAGTAGCTAAAATCTCCATGCGCGCTTAAACACA---CCCGGCCAGTTTGGAGAT 949
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
163 SerLeuArgIleThrSerLeuLeuArgThrValSerThrMetArgGlyGlyTyValPro 182
QY 950 TTCTTCCCGGAGCAGCGCAGACCAATCATCTCTTGCAGGGCTTCAGCAGGAATACG 1009
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
183 TyTyTyValAlaGlyArgAsnProGluThrValTySerAlaPheSerAspAspVal 202
QY 1010 TTGAGGCGCGCTTCAATCGGAATTCATAGATACGGAGG-----GTGCTGCTTA 1060
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
203 LeuGluAlaAlaPheAsnThrAsnValIleGluAlaArgThrHisPheProValHisIle 222
QY 1061 GAAGAGAATGTCAGGAGGTGAGCAAGAGAGAGAGGCGCAGAGGTGAGTACTCGGAGT 1120
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
223 GluArg-----GluSerTySer 228
QY 1121 AGTGAGAACATGAAGAGGTAGTAGTCAAGTGTCAAGAGGACGCTTGAAGAACTTACT 1180
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
229 MetAlaAsnGluGluGlnIle-----ArgGluMetLeuArg 240
QY 1181 AAGCAGCTAAAATCCGCTCAAGAAGAAGGCTCCGAAGAAGAGGAGGAGATATACCAACCCA 1240
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
241 LysArgGlyPheSerAlaGluSerMetSerAlaSerGluHis-----ProLysPro 257
QY 1241 ATCAACTTGAGAGAAGCGCGCGATCTTTCTAACAACCTTTGGGAAGTTTATTGAGGTG 1300
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
258 PheAsnLeuArgAsnGlnLysProAspPheGluAsnAspAsnGlyArgPheThrArgAla 277
QY 1301 AAGCCAGACAAGAAAGAACCCCGAGCTTCAGGACCTGACACATGATGCTACCTGTGTAGAG 1360
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
278 GlyProAsn---GluAsnProLeuLeuAspAlaValAspValThrAlaGlyPheGlyVal 296
QY 1361 ATCAAGAAGGAGCTTTGATGCTCCACACTTCAACTCAAGGCCCATGGTTATCGTCGTC 1420
Db   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
297 LeuAsnProGlyThrMetThrAlaProSerHisAsnThrLysAlaThrSerIleAlaIle 316
```



```
Db 274 Phe----- 274
Qy 1115 CGGAGTAGTGAGAACATGAGAGTGTATGCTCAAGTGTCAAAGGAGCACGTTGTGAAGAA 1174
Db 275 -----AspGlnGlnAsnGluGlySerIlePheArgIleSerArgGluGlnValArgAla 292
Qy 1175 CTTACTAAGCAGCGTAATCGGTCCTCAAGAAAGGCTCC-----GAAGAAGAGGGAGAT 1228
Db 293 Leu-----AlaProThrLysLysSerSerTrpTrpPheGlyGlyGlu 307
Qy 1229 ATCACCAACCAATCAACTTGAGAGAAGCGGAGCGGATCTTTCTAACAACCTTTGGGAAG 1288
Db 308 SerLysProGlnPheAsnIlePheSerLysArgProThrIleSerAsnGlyTyrGlyArg 327
Qy 1289 TTATTGAGGTGAAGCCA--GACAAGAAGAACCCCGAGCTTCAGGACCTCGACATGATG 1345
Db 328 LeuThrGluValGlyProAspAspGluLysSerTrpLeuGlnArgLeuAsnLeuMet 347
Qy 1346 CTCACCTGTGTAGAGATCAAGAAGAGGAGCTTTGATGCTCCACACACTTCAACTCAAGGCC 1405
Db 348 LeuThrPheThrAsnIleThrGlnArgSerMetSerThrIleHisTyrAsnSerHisAla 367
Qy 1406 ATGTTTATCGTCGTCAACAAGGAACCTGGAAACCTTGAACCTCGTGGCTGTAAGAAAA 1465
Db 368 ThrLysIleAlaLeuValIleAspGlyArgGlyHisLeuGlnIleSerCysProHisMet 387
Qy 1466 GAGCAACAACAGAGGGGACCGCGGAAGAGAGGAGGACGAGACGAAAGACGAAGAAGAGAGGGA 1525
Db 388 SerSerArgSerSerHisSerLysHisAspLys----- 398
Qy 1526 AGTAACAGAGAGGTGCGTAGGTACACAGCGAGGTGAAGAGCGGATGTTCATCATG 1585
Db 399 SerSerProSerTyrHisArgIleSerSerAspLeuLysProGlyMetValPheValVal 418
Qy 1586 CCAGCAGCTCATCCAGTAGCCATCAACGCTTCCTCCGAACCTCCATCTGCTT-----GGC 1639
Db 419 ProProGlyHisProPheValThrIleAlaSerAsnLysGluAsnLeuLeuMetIleCys 438
Qy 1640 TTCGGTATCAACGCT 1654
Db 439 PheGluValAsnAla 443
```

Search completed: April 8, 2005, 05:34:08

Job time : 104.925 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 8, 2005, 04:05:21 ; Search time 68.9539 Seconds

(without alignments)
5670.817 Million cell updates/sec

Title: US-10-728-051-1

Perfect score: 3562

Sequence: 1 aataatcatatattatcatc.....cgttgtgcgtgtttctcc 2032

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2_1/USPTO.spool_p/US10728051/runat_07042005_125045_14060/app_query.fasta_1.4757
-DB=PIR -QPM=fastan -SUFFIX=rdx -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10728051@cgn 1 1 228 @runat_07042005_125045_14060 -NCPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR_79: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1282	36.0	605	2	S20007
2	1281	36.0	605	1	FWSYBA
3	1267	35.6	571	2	S00566
4	1251	35.1	639	2	B24810
5	1178.5	33.1	459	2	S08505
6	1174.5	33.0	433	2	S00567
7	1156	32.5	439	1	FWSVCB
8	1148	32.2	438	2	S35757
9	1109.5	31.1	463	2	A27288
10	1108.5	31.1	463	2	S06309
11	1068	30.0	410	1	FWPMBV
12	962.5	27.0	445	2	J02264
13	959.5	26.9	445	2	S00281
14	907.5	25.5	810	2	T44430

ALIGNMENTS

RESULT 1

S20007

beta-conglycinin alpha chain precursor - soybean

C;Species: Glycine max (soybean)

C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004

C;Accession: S20007

R;Lelievre, J.M.; Dickinson, C.D.; Dickinson, L.A.; Nielsen, N.C.

Plant Mol. Biol. 18, 259-274, 1992

A;Title: Synthesis and assembly of soybean beta-conglycinin in vitro.

A;Reference number: S20007; MUID:92119248; PMID:1731988

A;Accession: S20007

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-605 <LEL>

A;Cross-references: UNIPROT:Q94LX2

C;Superfamily: glycinin

Alignment Scores:

Pred. No.: 8.03e-82 Length: 605

Score: 1282.00 Matches: 285

Percent Similarity: 59.37% Conservative: 108

Best Local Similarity: 43.05% Mismatches: 173

Query Match: 35.99% Indels: 96

DB: 2 Gaps: 18

US-10-728-051-1 (1-2032) x S20007 (1-605)

QY	50	ATGACGAGGAGGGTCTTCTCCACGTAGTCTGTCTCTAGGATCCTGTCTGGCTTCAGTT	109
Db	2	MetArgAlaArgPhePro-----LeuLeuLeuGlyLeuValPheLeuAlaSerVal	19
QY	110	TCGTGCAACGATCCCAAGTCATCCTTACCAGAGAAACAGACACCTCGGCCAG	169
Db	20	SerValSerPheGlyIleAla-----TyrTrpGluLys---GluAsnProLysHisAsn	36
QY	170	AGTGTCTCCAGATTGTCAACAGGACCGATGACTTGAAGCAAGGAGGATCGAGTCT	229
Db	37	LysCysLeuGlnSerCysAsnSerGluArgAspSerTyrArgAsnGlnAlaCysHisAla	56

alpha-globulin typ
phaseolin type alp
vicilin precursor
alpha-phaseolin pr
phaseolin beta cha
alpha-globulin typ
alpha-globulin B p
vicilin A precursor
vicilin-like stora
globulin-1S, GUB1S
globulin Begi prec
beta-conglycinin a
globulin-2 precurs
62K sucrose-bindin
62K sucrose-bindin
vicilin-like stora
7S storage protein
convicilin (clone
vicilin-like stora
vicilin - zamia fu
convicilin precurs
phaseolin - kidney
probable major pro
vicilin, 14K compo
globulin1 - maize
globulin-1 - maize
hypothetical prote
probable seed stor
hypothetical prote
globulin-like prot
beta-globulin B pr

	CGCTGCACCAAGCTCCGAGTATCAT	-----CCTCGTTGT 262 :::
D _b	57 ArgCysAsnLeuLeuLysValGlulysGluGluCysGluGlyGluIleProArg---	----75
QY	263 GTCTATGATTCCTCGAGGACACACTGGCCACCACAACCAACCTGCCCTCCAGGGAGCGG	322
D _b	76 -----ProArgProArg----	84
QY	323 ACAGTGCCCCCACAAACCCGGAGACTACGATGATGAC	358 ::: ::: :::
D _b	85 ArgGluProGlnGlnProGlyGluLysGluAsePGLiuAspGluInProArgProfile	104
QY	359 -----CGCGGTCAAACCCCGGAAGACAG-	391
D _b	105 ProPheProArgProGlnProArgGlnGluHisGluGlnArgGluGluGlnGlu	124
QY	392 TGG	430
D _b	125 TrpProArgLysGluGluLysArgGlyGluLysGlySerGluGluGluAspGluaspGlu	144
QY	431 GACTGGAGAACCAAGAAGAAGAT-----TGGAGGGACCAAGTTCATCAGCACCCACGG	484
D _b	145 AspGluGluInAspGluUargGlnPheProPheProArgProHisGlnLysGluGlu	164
QY	485 AAAATAAGCCCAGGAAGAAGAGAGAACAAAGTAGTGGGAAACACAGGTAGCCATGTG	544 ::: ::: ::: ::: ::: :::
D _b	165 ArgyLysGlnGluGluAspGluaspluglulglnarGlsrGluSerGluUsrGluas	184
QY	545 AGGGAAA-----GAAAACTCCGGAACAAACCCCTTTCTACTTCCCGTCAAGCGCGTTTTAGC	598
D _b	185 SerGluLeuArgArGHisLysAsnLysAsnProPheLeuPheGlySerAsnarGrPheGlu	204
QY	599 ACCCGTACTGGGAACCAAAACGGTAGGATCCGGGTCTCTGCAGAGTTGTACCAAAGTCA	658 :::
D _b	205 ThrLeuPheLysAsenGlnTyrgLYarginileArgVallLeuGlnArGPHeasnGlnArgSer	224
QY	659 AGGCAGATTTCAGAACTCCAGAACTCCGATTAATCTGCAGATCGAGGCCCAACCTTAACACT	718
D _b	225 ProGlnLeuGlnAsnLeuUrGaspyrArgilLeuGluPheAsnSerLySPrOasnThr	244
QY	719 CTGTCTTCTCCAAGCAGCTGATGCTGATAACATCCTTGTTATCCAGCAAGGGCAAGCC	778
D _b	245 LeuLeuLeuProAsnHisaAlaaSpTYrlleulleValllleLeuasnglyThraLa	264
QY	779 ACCGTGACGTAGCAAAATGGCAATAACAGAAAGACTTTAATCTTGACGAGGGCCATGCA	838
D _b	265 IleLeuserrLeualAsnaAsnapSParspserTyrArgLeuGlnSerGIyaspala	284
QY	839 CTCAGAAATCCCATCCCGTTTCATTTCTACATCTTGAACCCCNATGACACACCAACCTC	898
D _b	285 LeuarGalProSerGlyThrThrytyrvalValAsnPProaspAEnsnGluasnLeu	304
QY	899 AGAGTAGCTAAAFCTCCATGCCCGTTTAAACACACCCCGCCAGTTTGAGGATTTCTCCCG	958
D _b	305 ArgLeulethrLeualaIleProvaIsanLysProgIyarPherGluSerPhePheLeu	324
QY	959 CGGAGCAGCCGAGACCAATCATCTACTTCGAGGGCTTCACGAGGAATACGTTGGAGGCC	1018 ::: ::: ::: ::: :::
D _b	325 SerSerThrGluAlaGlnInSerTyrLeuGlnglyPheSerArgAsnilleLeuGluAla	344
QY	1019 GCCTTCAATGCGGAATTCATATGATACGAGGGTGTCTGTAGAGAGAATGCAGGAGGT	1078 ::: ::: ::: ::: :::
D _b	345 SerTyraspThryspPheGluGluIlEasnLysvallLeuPheSerArgGluGlugly--	363
QY	1079 GAGCAAGAGGAGAGAGGCGCAGCGCATGCGAGTACTCGGAGTAGTGAGAACAAATGAGGA	1138
D _b	364 -----GlngInglinglyGluGlnArgLeu-----GingLuser	374
QY	1139 GTGNATGCTAAAGTCTCAAAGGACACGTTGAAGNACTTACTAAGCACGCTAAATCCGCT	1198
D _b	375 ValIleValGluIleSerLysGluGlnIleArgalaleuSerLysArglaLalySerSer	394

QY	1199	TCAAGAAAGGCTCCGAGAGAGAGAGGAGATATACCAACCCATCAACTTACACTTACAGAGAAGGC	1258
Db	395	SerArglystThrIleSerSerGlu-----AspLysProPheAsnLeuArgSerArg	411
QY	1259	GAGCCCGACTTTCTAAACAATTGGGAAGTTATTTGAGCTGAAGCCAGACAAGAAGAAC	1318
Db	412	AspProIleTyrSerAsnLysLeuGlyLysPhePheGluIleThrProGlu---LysAsn	430
QY	1319	CCCAGCTTCAGGACCTGGACATGATGCTCACCTGTGTAGAGATCAAGAAGAGAGCTTTG	1378
Db	431	ProGlnLeuArgAspLeuAspIlePheLeuSerIleValAspMetAsnGluGlyAlaLeu	450
QY	1379	ATGCTCCCACTTCAACTCAAGGCCATCGTTATCGTCGCTCAACAAGAACTCGGA	1438
Db	451	LeuLeuProHisPheAsnSerLysAlaIleValIleLeuValIleAsnGluGlyAspAla	470
QY	1439	AACCTTGAACCTCGTGGCTGTAAAGAAAGACAAACAAGAGGGGACGGCGGGAAGAAGAG	1498
Db	471	AsnIleGluLeuValGlyLeu---LysGluGlnGlnGlnGlnGlnGlnGluGln	489
QY	1499	GAGGAAGACGACGAAGAGAGGGGAAGTAAACAGAGAGGTGCGTAGGTACACAGCGAGG	1558
Db	490	Pro-----LeuGluValArgLysTyrArgAlaGlu	499
QY	1559	TTGAAGGAAGCGATGTTTCATCATGCCAGCAGCTCATCCAGTAGCCATCAACGCTTCC	1618
Db	500	LeuSerGluGlnAspIlePheValIleProAlaGlyTyrProValValValAsnAlaThr	519
QY	1619	TCCGAACCTCCATCTGCTTGGCTTCGCTATCAACGCTGAAACACACACAGAACTCTTCC	1678
Db	520	SerAsnLeuAsnPhePheAlaIleGlyIleAsnAlaGluAsnGlnArgAsnPheLeu	539
QY	1679	GCAGGTGATTAAGCAATGTGATAGACCAGATAGAGAAGCAACGCAAGGATTTAGCATTC	1738
Db	540	AlaGlySerGlnAspAsnValIleSerGlnIleProSerGlnValGlnGluLeuAlaPhe	559
QY	1739	CCTGGGTCGGGTGAACAAAGTTGAGAAGCTCATCAAAAAACAGAAGGAATCTCACTTTG	1798
Db	560	ProGlySerAlaGlnAlaValGluLysLeuLeuLysAsnGlnArgGluSerTyrPheVal	579
QY	1799	AGTGTCTGCTCAATCTCAATCTCAATCTCGCTCTCTCTCTGAGAAAGAGTCTCTGAG	1858
Db	580	AspAlaGlnProLysLys-----	585
QY	1859	AAAGAGGATCAAGAGGAGGAAACCAGAGGGAAGGCTCCACTCTCTTTCAATTTTGAAG	1918
Db	586	-----LysGluGluGlyAsnLysGlyArgLysGlyProLeuSerSerIleLeuArg	602
QY	1919	GCATTTT	1924
Db	603	AlaPhe	604

RESULT 2

FWSYBA

beta-conglycinin alpha chain precursor - soybean

C:Species: Glycine max (soybean)

C:Date: 31-Dec-1991 #sequence revision 31-Dec-1991 #text_change 09-Jul-2004

C:Accession: S14681; S74124; S06714

R:Sebastiani, P.L.; Farrell, L.B.; Schuler, M.A.; Beachy, R.N.

P:Plant Mol. Biol. 15, 197-201, 1990

A:Title: Complete sequence of a cDNA of alpha subunit of soybean beta-conglycinin

A:Reference number: S14681; PMID:91355860; PMID:2103438

RESULT 2

RESUL
FWSYBA

FWSIDA
beta-conglucinin alpha chain precursor - carbox

beta-conglutinin antigen (see below).

C;species: Glycine max (soybean)

```
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_chang
C;sequence: 014601-074124: 006714
```

C;Accession: S14681; S74124; S06714

R;Sebastiani, F.L.; Farrell, L.B.; Schuler, M.A.; Beachy, R.N.

Plant Mol. Biol. 15, 197-201, 1990

A;Title: Complete sequence of a cDNA of alpha subunit of soybean beta-conglycinin.

A;Reference number: S14681; MUID:91355860; PMID:2103438

A;Accession: S14681

A;Molecule type: mRNA

A;Residues: 1-605 <SEB>

A;Cross-references: UNIPROT:P13916; EMBL:X17698; NID:g18535; PIDN:CAA35691.1; PID:g18536

R; Shutov, A. D.; Kakhovskaya, I. A.;

Eur. J. Biochem. 241, 221-228, 1996

A;Title: Limited proteolysis of beta-conglycinin and glycinin, the 7S and 11S storage glc

A;Reference number: S74123; MUID:97054613; PMID:8898910

A;Accession: S74124

A;Molecule type: protein

A;Residues: 189-196, 'H', 198, 'N', 200, 'X', 202-203;397-408, 'X', 410, 'X', 412-417, 'X', 419-420, '

A: Experimental source: seed

C: Superfamily: glycinin

F: 1-22/Domain: signal sequence #status predicted <SIG>

F: 23-62/Domain: propeptide #status predicted <PRO>

F: 63-605/Product: beta-conglycinin alpha chain #status predicted <WAT>

F: 261,517/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:

Pred. No.:	9,43e-82	Length:	605
Score:	1281.00	Matches:	285
Percent Similarity:	59.21%	Conservative:	107
Best Local Similarity:	43.05%	Mismatches:	174
Query Match:	35.96%	Indels:	96
DB:	1	Gaps:	18

US-10-728-051-1 (1-2032) x FWSYBA (1-605)

Qy	50	ATGAGGGAGGTTTCTCCACTGATCTGTTGCTAGGATCTTGTCTGGCTTCAGTT	109
Db	2	MetArgAlaArgPhePro-----LeuLeuLeuLeuGlyLeuValPheLeuAlaSerVal	19
Qy	110	TCTCAACGGATGCCAAGTCATCCTTACCAGAGAAACAGAGAACCCTCGGCCAG	169
Db	20	SerValSerPheGlyIleAla-----TyrTrpGluLys---GluAsnProLysHisAsn	36
Qy	170	AGTGCTCCAGAGTTGTCACAGAAACCGGATGACTTTGAAGCAAAAGGATCGAGTCT	229
Db	37	LysCysLeuGlnSerCysAsnSerGluArgAspSerTyrArgAsnGlnAlaCysHisAla	56
Qy	230	CGTGCACCAAGCTCGAGTATGAT-----CCTCGTTGT	262
Db	57	ArgCysAsnLeuLeuLysValGluLysGluCysGluGluGlyGluIleProArg---	75
Qy	263	GTCTATGATCTCCAGGACACACTGGCACCACCAACCAACGTTCCCTCCAGGGACGG	322
Db	76	-----ProArgProArg-----ProGlnHisProGlu	84
Qy	323	ACAGTGGCGGCCCAACCCGGAGACTAGCATGATGAC-----	358
Db	85	ArgGluProGlnGlnProGlyGluLysGluAspGluAspGluGlnProArgProIle	104
Qy	359	-----CGCCGTCAACCCCGAAGAG-----GAAGAGGCCGA	391
Db	105	ProPheProArgProGlnProArgGlnGluGluHisGluGlnArgGluGluGlnGlu	124
Qy	392	TGG-----GGACCAGCTGGACGGAGGCGTGAAGAGAGAA	430
Db	125	TrpProArgLysGluGluLysArgGlyGlyLysGlySerGluGluGluAspGlu	144
Qy	431	GACTGGAGACAACCAAGAGAAGAT-----TGGAGGCGACCAAGTCATCAGCCACGG	484
Db	145	AspGluGlnAspGluArgGlnPheProArgProHisGlnLysGluGlu	164
Qy	485	AAATAAGGCCGGAAGAGAGAGAGACAAGATGGGGAACACACGAGTAGCCATGTG	544
Db	165	ArgAsnGluGluGluAspGluAspGluGlnGlnArgGluSerGluSerGluAsp	184
Qy	545	AGGAA-----GAAACATCTCGAAACACCTTTCTACTCCCGTCAAGCGGTTTAGC	598
Db	185	SerGluLeuArgArgHisLysAsnLysAsnProPheLeuPheGlySerAsnArgPheGlu	204
Qy	599	ACCGCTACGGGAACCAACGGTAGGATCCGGTCTCGAGAGGTTTGAACCAAGTCA	658
Db	205	ThrLeuPheLysAsnGlnTyrGlyIleArgValLeuGlnArgPheAsnGlnArgSer	224
Qy	659	AGGCAGTTTCAGAATCTCCGAATCACCCTATTTCGAGATCCGAGCCAAACCTTAACACT	718
Db	225	ProGlnLeuGlnAsnLeuArgAspTyrArgIleLeuGluPheAsnSerLysProAsnThr	244
Qy	719	CTTGTTCTTCCAGGACGCTGATGATACCTCTGTTATCCAGCAAGGCGCAAGCC	778
Db	245	LeuLeuLeuProAsnHisAlaAspAlaAspTyrLeuIleLeuAsnGlyThrAla	264

Qy	779	ACCGTGACCGTAGCAATGCAATACAGAAAGAGCTTTAATCTTGACGAGGGCCATGCA	838
Db	265	IleLeuSerLeuValAsnAsnAspArgAspSerTyrArgLeuGlnSerGlyAspAla	284
Qy	839	CTCAGATCCCATCCGGTTTCATTCCTACATCTTGAAACCGCATGACCAACCACTC	898
Db	285	LeuArgValProSerGlyThrThrTyrTyrValValAsnProAspAsnAsnGluLeu	304
Qy	899	AGAGTAGTAAATCTCCATGCCCGCTTAAACACACCCCGCCAGTTTGAAGGATTTCTCCCG	958
Db	305	ArgLeuIleThrLeuAlaIleProValAsnLysProGlyArgPheGluSerPhePheLeu	324
Qy	959	GCGAGCAGCGAGACCAATCATCTCTTCCAGGGCTTACAGCAGGAATATGTTGGAGGCC	1018
Db	325	SerSerThrGluAlaGlnGlnSerTyrLeuGlnGlyPheSerArgAsnIleLeuGluAla	344
Qy	1019	GCCTTCAATCGGGAATTCATGATGATACGAGGGTCTGTTAGNAGNAGTGCAGGAGGT	1078
Db	345	SerTyrAspThrLysPheGluGluIleAsnLysValLeuPheSerArgGluGluGly---	363
Qy	1079	GAGCAAGAGGAGAGGAGCGAGCGATGAGTACTCGGAGTAGTGAGAACAAATGAAGGA	1138
Db	364	-----GlnGlnGlnGlyGlnArgLeu-----GlnGluSer	374
Qy	1139	GTGATAGTCAAGTGTCAAAGGAGCAGTTTGAAGAACTTACTTAAGCAGCTTAATCCGTC	1198
Db	375	ValIleValGluIleSerLysGluGlnIleArgAlaLeuSerLysArgAlaLysSerSer	394
Qy	1199	TCAAAGAAAGGCTCCGAGAGAGGAGATATCACCACCAATCACTTGAAGAGAGGC	1258
Db	395	SerArgLysThrIleSerSerGlu-----AspLysProPheAsnLeuArgSerArg	411
Qy	1259	GAGCCGATCTTCTAACTTTGGAGTATTTGAGGTGAAGCCAGCAACAAGAAC	1318
Db	412	AspProIleTyrSerAsnLysLeuGlyLysPheGluIleThrProGlu---LysAsn	430
Qy	1319	CCCAGCTTCAGACCTGGACATGCTCCTCCTGTGTAGATCAAAAGAGGAGCTTTG	1378
Db	431	ProGlnLeuArgAspLeuAspIlePheLeuSerIleValAspMetAsnGluGlyAlaLeu	450
Qy	1379	ATGCTCCACACTCAACTCAAGGCCATGTTATCGTCTGTCATCAAAAGGAGCTTGA	1438
Db	451	LeuLeuProHisPheAsnSerLysAlaIleValIleLeuValIleAsnGluGlyAspAla	470
Qy	1439	AACCTTGAACTCTGGCTGTAAGAAAGACAACAAGAGGCGACGCGGGAAGAGAG	1498
Db	471	AsnIleGluLeuValGlyLeu---LysGluGlnGlnGlnGlnGlnGlnGluGln	489
Qy	1499	GAGGACGAAGACGAAGAGGAGGGAAGTAACAGAGAGGTGCTAGTACACAGCGAGG	1558
Db	490	Pro-----LeuGluValArgLysTyrArgAlaGlu	499
Qy	1559	TTGAAGAGGCGATGTTTCATGTCAGAGCTCATCCAGTAGCCATCAACGCTTCC	1618
Db	500	LeuSerGluGlnAspIlePheValIleProAlaGlyTyrProValValAsnAlaThr	519
Qy	1619	TCCGAACTCCATCTGCTGCTTCCGTTATCAAGCTGAAACCAACCAAGATCTTCTT	1678
Db	520	SerAsnLeuAsnPhePheAlaIleGlyIleAsnAlaGluAsnAsnGlnArgAsnPheLeu	539
Qy	1679	GCAGGTGATAGGACAATGTGATAGACAGATAGAGAAAGCAAGCAAGGATTTAGCATTC	1738
Db	540	AlaGlySerGlnAspAsnValIleSerGlnIleProSerGlnValGlnGluLeuAlaPhe	559
Qy	1739	CTTGGGTGCGGTGAACAAGTTGAGAGCTCATCAAAACCAAGAGGAATCTCATTTGTG	1798
Db	560	ProGlySerAlaAlaValGluLysLeuLeuLysAsnGlnArgGluSerTyrPheVal	579
Qy	1799	AGTGCTGCTCTCAATCTCAATCTCCGTCGCTCTCTGAGAAAGAGTCTCTCTGAG	1858
Db	580	AspAlaGlnProLysLys-----	585

```
QY 1859 AAAGAGGATCAAGAGGAGGAAAACCAAGGAGGAGGTCCTCACTCTTTCAATTTTGAAG 1918
Db 586 -----LysGluGluGlyAsnLysGlyArgLysGlyProLeuSerSerIleLeuArg 602
QY 1919 GCTTTT 1924
Db 603 AlaPhe 604

RESULT 3
S00566
convicillin precursor (clone pJC4) - garden pea
C:Species: Pisum sativum (garden pea)
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 09-Jul-2004
C:Accession: S00566
R:Brown, D.; Ellis, T.H.N.; Gatehouse, J.A.
Biochem. J. 251, 717-726, 1988
A:Title: The sequence of a gene encoding convicillin from pea (Pisum sativum L.) shows th
A:Reference number: S00566; MUID:88326208; PMID:3415641
A:Accession: S00566
A:Molecule type: DNA
A:Residues: 1-571 <BOW>
A:Cross-references: UNIPROT:P13915; EMBL:X06398; NID:q20698; PIDN:CAA29695.1; PID:g31367
A:Note: part of this sequence, including the amino end of the mature protein, was confir
A:Note: 237-Gln was also found
C:Genetics:
A:Gene: cvcA
A:Map position: 2
A:Introns: 221/1; 279/3; 304/3; 412/3; 507/1
C:Superfamily: glycine
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-571/Product: convicillin #status experimental <MAT>

Alignment Scores:
Pred. No.: 8.9e-81 Length: 571
Score: 1267.00 Matches: 276
Percent Similarity: 56.67% Conservative: 98
Best Local Similarity: 41.82% Mismatches: 146
Query Match: 35.57% Indels: 140
DB: 2 Gaps: 13

US-10-728-051-1 (1-2032) x S00566 (1-571)

QY 68 CCACTGATGCTGTGTAGGATCTTGTCTCGCTTCTGCTTCTGCAAGCATGCCAAG 127
Db 10 ProLeuLeuPheLeuGlyIlePheLeuAlaSerValThrTyAlaAsn 29
QY 128 TCATCCTTACCAGAAGAAAACAGAACCCCTGCGCCGAGGTGCTCCAGAGTTGT 187
Db 29 ----- 29
QY 188 CAACAGGAACCGGATGATTTGAAGCAAAAGGATGCGAGTCTCGCTGCCAACCAAGCTCGAG 247
Db 29 ----- 29
QY 248 TATGATCCTGCTGTCTATGATCTCTCGAGGACACACTGGCACCAACCAACCAAGTTCC 307
Db 30 -----TyrAspGluGlySerGluThrArgVal----- 38
QY 308 CCTCCAGGGAGCGACACGTGGCCGCCAACCCGGAGACTACCATGATGACCGCCGTCAA 367
Db 39 -----ProGlyArgGluArgGlyArgGlnGluGlyGluGlyArgHis----- 56
QY 368 CCCCGAAGAGAGGAGCGGATGGGACCGACTGGACCGGAGGAGCGT----- 418
Db 57 -----GlyGluTrpArgProSerTyrGluLysGluGluHisGluGlu 70
QY 418 ----- 418
Db 71 GluLysGlnLysTyrArgTyrGlnArgGluLysGluGlnLysGluValGlnProGly 90
QY 419 -----GAAAGAGAGAGACTGGAGACCAACCAAGAGAGATTGGAGG----- 460
Db 91 ArgGluArgTrpGluArgGluGluAspGluGluGlnValGluGluGluTrpArgGlySer 110
```

```
QY 461 -----CGACCAAGTCATCAGCAGCCACGG 484
Db 111 GlnArgArgGluAspProGluGluArgAlaArgLeuArgHisArgGluGluArgThrLys 130
QY 485 AAAATAAGCCCGAAGGAAAGAGAGAAAGAGTGGGGAACACCAAGGTAGCCATGTG 544
Db 131 ArgAspArgArgHisGlnArgGluGlyGluGlu-----GluArgSer 145
QY 545 AGGGAAGAAACATCTCGGNACCAACCTTCTACTTCCCGTCAGGCGGTTTAGCCCGC 604
Db 146 SerGluSerGlnGluHisArgAsnProPheLeuPheLysSerAsnLysPheLeuThrLeu 165
QY 605 TAGCGGAACCAAAACCGTAGGATCCGGTCTCGACAGAGTTTGACCAAAAGGTCAAGGCAG 664
Db 166 PheGluAsnGluAsnGlyHisIleArgArgLeuGlnArgPheAspLysArgSerAspLeu 185
QY 665 TTTTCAAGATCTCCAGATCAACCGTATTGTGAGATGAGGCCAAACCTAACACTCTTGT 724
Db 186 PheGluAsnLeuGlnAsnTyrArgLeuValGluTyrArgAlaLysProHisThrIlePhe 205
QY 725 CTTCCCAAGCAGCTGATGCTGATACATCTTGTATTATCCAGCAAGGGCAAGCACCGTG 784
Db 206 LeuProGlnHisIleAspAlaAspLeuIleValLeuAsnGlyLysAlaIleLeu 225
QY 785 ACCGTAGCAATGGCAATAACAGAAAGAGCTTTAATCTTTCAGGAGGCCATGCACCTCAGA 844
Db 226 ThrValLeuSerProAsnAspArgAsnSerTyrAsnLeuGluArgGlyAspThrIleLys 245
QY 845 ATCCCATCCGGTTTCATTTCTTCTTGAACCGCATGACACACAGAACCTCAGAGTA 904
Db 246 IleProAlaGlyThrThrSerTyrLeuValAsnGlnAspGluGluAspLeuArgVal 265
QY 905 GCTAAATCTCATGCCGTAAACACACCGCCAGTTTGTAGGATTTCTCCGCGGAGC 964
Db 266 ValAspPheValIleProValAsnArgProGlyLysPheGluAlaPhe-----GlyLeu 283
QY 965 AGCCGAGACCAATCATCTTTCAGGAGCTTTCAGCAGGAATACGTTGAGGCGCGCTTC 1024
Db 284 SerGluAsnLysAsnGlnTyrLeuArgGlyPheSerLysAsnIleLeuGluAlaSerLeu 303
QY 1025 AATGGGAATTCATAGATACGGAGGTGCTGTAGAGAGATGCCAGAGGTGAGCA 1084
Db 304 AsnThrLysTyrGluThrIleGluLysValLeuLeuGluGlnGlnLysLysProGln 323
QY 1085 GAGGAGAGGAGCGAGCGCATGAGTACTCGAGTAGTGAGAACAAATCAAGAGAGTGATA 1144
Db 324 GlnLeuArgAspArgLysArg-----ThrGlnGlnGlyGluGluArgAsp---AlaIle 340
QY 1145 GTCAAAGTCTCAAAGGAGCAGCTTGAAGAACTTACTAAGCAGCTTAATCCGTCTCAAAG 1204
Db 341 IleLysValSerArgGluGlnIleGluGluLeuArgLysLeuAlaLysSerSerLys 360
QY 1205 AAGGCTCCGAGAGAGGAGATATCACCAACCAATCAACTTGAAGAGGCGGAGCC 1264
Db 361 LysSerLeuProSerGluPheGlu-----ProPheAsnLeuArgSerHisLysPro 377
QY 1265 GATCTTTCTAAACACTTTGGGAAGTTATTGTAGGTGAAGCCAGACAAAGAACCCCGCAG 1324
Db 378 GluTyrSerAsnLysPheGlyLysLeuPheGluIleThrProGluLysLysTyrProGln 397
QY 1325 CTTTCAAGACTCGACATGATGCTACCTGTGTAGATCAAGAGAGACTTTGATGCTC 1384
Db 398 LeuGlnAspLeuAspIleLeuValSerCysValGluIleAsnLysGlyAlaLeuMetLeu 417
QY 1385 CCACACTTCAACTCAAAGGCCATGTTATCGTCGTCTCAACAAAGAACTCGAAACCTT 1444
Db 418 ProHisTyrAsnSerArgAlaIleValIleValLeuValAsnGluGlyLysGlyAsnLeu 437
QY 1445 GAACCTCGCTGTAGAAAAGAGCAACACAGAGGGGCGCGGGAAGAGAGGAGAC 1504
Db 438 GluLeuLeuGlyLeuLysAsnGlnGlnGluArg----- 449
```

```
QY 1505 GAAGACGAAGAAGAGGAGGAAGTAACAGAGAGGTGCGTAGGTACACAGCAGGCTTGAAG 1564
Db 450 GluAspArgLysGluArg-----AsnAsnGluValGlnArgTyrGluAlaArgLeuSer 467
QY 1565 GAAGGCGATGTTTCATCATGCCAGCAGCTCATCCAGTACGATCCATCAACGCTTCTCCGAA 1624
Db 468 ProGlyAspValValIleIleProAlaGlyHisProValAlaIleSerAlaSerAsn 487
QY 1625 CTCATCTGCTTGGCTTCGGTATCAACGCTGAAACACACAGAAATCTTCTTGCAGGT 1684
Db 488 LeuAsnLeuLeuGlyPheGlyIleAsnAlaLysAsnAsnGlnArgAsnPheLeuSerGly 507
QY 1685 GATAAGACATGTATGATACCATAGACATAGACAGCAACGCAAGATTTAGCATTCCTCCGG 1744
Db 508 SerAspAsnValIleSerGlnIleGluAsnProValLysGluLeuThrPheProGly 527
QY 1745 TCGGTCAACAAGTTGAGAAGCTCATCAAAACAGAAAGAAATCTCACTTTGTGAGTGCT 1804
Db 528 SerSerGlnGluValAsnArgLeuIleLysAsnGlnLysGlnSerHisPheAlaSerAla 547
QY 1805 CGTCTCAATCTCAATCTCAATCTCCGCTGCTCTGAGAAAGAGTCTCTGAGAAAGAG 1864
Db 548 GluPro-----Glu 550
QY 1865 GATCAAGAGAGGAAACCAAGGAGGAGGAGGTCCTCTTCAATTTGAAGGCTTTT 1924
Db 551 GlnLysGluGluGluSerGlnArgLysArgSerProLeuSerSerValLeuAspSerPhe 570

RESULT 4
B24810
beta-conglycinin alpha chain precursor - soybean
C:Species: Glycine max (soybean)
C>Date: 07-Mar-1988 #sequence_revision 30-Jun-1991 #text_change 09-Jul-2004
C:Accession: B24810; MUID:86250867; PMID:3013879
R:Doyle, J.J.; Schuler, M.A.; Godette, W.D.; Zenger, V.; Beachy, R.N.; Slightom, J.L.
J. Biol. Chem. 261, 9228-9238, 1986
A:Title: The glycosylated seed storage proteins of Glycine max and Phaseolus vulgaris. S
A:Reference number: A24810; MUID:86250867; PMID:3013879
A:Accession: B24810
A:Molecule type: DNA
A:Residues: 1-639 <DOY>
A:Cross-references: UNIPROT:P11827; GB:M13759; NID:g169928; PIDN:AAB01374.1; PID:g169928
R:Schuler, M.A.; Schmitt, E.S.; Beachy, R.N.
Nucleic Acids Res. 10, 8225-8244, 1982
A:Title: Closely related families of genes code for the alpha and alpha' subunits of the
A:Reference number: S16337; MUID:83143288; PMID:6298713
A:Accession: S16337
A:Molecule type: DNA
A:Residues: 361-639 <SCH>
A:Introns: 278/1; 355/3; 382/3; 481/3; 575/1
C:Superfamily: glycinin
C:Keywords: seed; storage protein

Alignment Scores:
Pred. No.: 1,26-79 Length: 639
Score: 1251.00 Matches: 283
Percent Similarity: 58.75% Conservative: 120
Best Local Similarity: 41.25% Mismatches: 173
Query Match: 35.12% Indels: 110
DB: 2 Gaps: 21

US-10-728-051-1 (1-2032) x B24810 (1-639)
QY 50 ATGAGAGGAGGTTTCTCCATGATCTGTTGCTAGGATCCCTGTGCTCGGCTTCAGTT 109
Db 2 MetArgAlaArgPhePro-----LeuLeuLeuLeuGlyValValPheLeuAlaSerVal 19
QY 110 TCTCAACGATGCCAAGTCATCACTTACCAGAAAGAAACAGAGAACCCTCGGCCAG 169
Db 20 SerValSerPheGlyIleAla-----TyrTrpGluLys---GlnAsnProSerHisAsn 36
```

```
QY 170 AGTGCTCTCAGAGTTGTCAACAGGAACCGGATGACTTGAAGCAAAAGGATCGCATCT 229
Db 37 LysCysLeuArgSerCysAsnSerGluLysAspSerTyrArgAsnGlnAlaCysHisAla 56
QY 230 CGCTGCGACC-----AAGCTCGAGTATGATCCTCGTTGT-----GTCTATGAT 271
Db 57 ArgCysAsnLeuLeuLysValGluGluGluGluCysGluGluGlnIleProArg 76
QY 272 CCTCGA---GGACACACTGCGCACCAACCAACCAACCTCCCTCCAGGCGGCGACACGT 328
Db 77 ProArgProGlnHisProGluArgGluArgGlnGlnHisGlyGluLysGluGluAspGlu 96
QY 329 GCGCGCAACCCGAGACTAGATGATACCGC---CGTCAACCCCAAGAGAGGAA--- 382
Db 97 GlyGluGlnProArgProPheProPheProArgProArgGlnProHisGlnGluGlu 116
QY 383 -----GGAGGCCGATGG 394
Db 117 HisGluGlnLysGluGluHisGluTrpHisArgLysGluLysHisGlyGlyLysGly 136
QY 395 GGA-----CCAGCTGCGACCGGAGGAGCGTGAA 421
Db 137 SerGluGluGluGlnAspGluArgGluHisProArgProHisGlnProHisGlnLysGlu 156
QY 422 AGAGAAGAA---GACTGGAGACAACCAAGAGAAGATTGGAGGCGACCAAGTCATCAGCAG 478
Db 157 GluGluLysHisGluTrpGlnHisLysGlnGluLysHisGlnGlyLysGluSerGluGlu 176
QY 479 CCAGGAAAATAAGGCCCGAAGGAGAGAGGAGAGCAACAG---TGGGGAACACCAAGGT 535
Db 177 GluGluGluAspGlnAspGluAspGluGluGluAspLysGluSerGlnGluSerGluGly 196
QY 536 AGCATGTGAGGGAAGAAACATCTCGGAAC-----AACCTTTTCTACTTCCCTCA 586
Db 197 SerGluSerGlnArgGluProArgArgHisLysAsnLysAsnProPheHisPheAsnSer 216
QY 587 AGCGGTTTAGACCCCGCTACGGAGAACCAAAACGGTAGATCCGGTCTCGCAGAGGTTT 646
Db 217 LysArgPheGlnThrLeuPheLysAsnGlnTyrGlyHisValArgValLeuGlnArgPhe 236
QY 647 GACCAAGGTCAAGGAGGTTTCAGATCTCCAGATCCACCGTATTGTGCAGATCGAGGCC 706
Db 237 AsnLysArgSerGlnGlnLeuGlnAsnLeuArgAspTyrArgIleLeuGluPheAsnSer 256
QY 707 AAACCTAAACACTCTTGTCTTCCCAAGCAGCTGATGATCAATCACTTGTATTCCAG 766
Db 257 LysProAsnThrLeuLeuLeuProHisHisAlaAspAlaAspTyrLeuIleValIleLeu 276
QY 767 CAAGGCGACGACCCGTCGACCGTAGCAATGGCAATACAGAAAGACGTTTATCTTGAC 826
Db 277 AsnGlyThrAlaIleLeuThrLeuValAsnAsnAspAspArgAspSerTyrAsnLeuGln 296
QY 827 GAGGGCCATGCTCAGATCCCATCCGTTTCTATTCTTACATCTTGAC----- 877
Db 297 SerGlyAspAlaLeuArgValProAlaGlyThrThrPheTyrValValAsnProAspAsn 316
QY 878 -----CGCCATGACAAC--- 889
Db 317 AspGluAsnLeuArgMetIleAlaGlyThrThrPheTyrValValAsnProAspAsnAsp 336
QY 890 CAGAACCTCAGATAGCTAAATCTCCATGCGCGTTTAAACACACCCGCGCAGTTGAGAT 949
Db 337 GluAsnLeuArgMetIleThrLeuAlaIleProValAsnLysProGlyArgPheGluSer 356
QY 950 TTCTTCCGCGAGCAGCCGAGACCATCATCTCTTGTGAGGCTTCACGGAATACG 1009
Db 357 PhePheLeuSerSerThrGlnAlaGlnGlnSerTyrLeuGlnGlyPheSerLysAsnIle 376
QY 1010 TTGGAGCCCGCTTCAATGCGGAATCAATGAGATCGAGGCTGTCTGTAGAGAGAAAT 1069
Db 377 LeuGluAlaSerTyrAspThrLysPheGluGluIleAsnLysValLeuPheGlyArgGlu 396
QY 1070 GCAGGAGGTGAGCAA---GAGGAGAGAGGCGAGCGGCTACTCGGAGTAGTGAG 1126
```



```
Db 300 LeuProHisTyrAsnSerArgAlaIleValIleValThrValAsnGluGlyLysGlyAsp 319
Qy 1442 CTTGAACCTCGTGTGTAAGAAAGAGCAACAACAGAGGGAGCGCGGGAAGAGGAG 1501
Db 320 PheGluLeuValGlyGlnArgAsnGluAsnGlnGlnGluGlnArgLysGluAspGlu 339
Qy 1502 GACGACACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1561
Db 340 GluGluGlnGlnGlyGluGluGluIleAsnLysGlnValGlnAsnTyrLysAlaLysLeu 359
Qy 1562 AAGGAGGCGCATGTTTCATCATCCGACGAGCTCATCCAGTAGCATCAACGCTTCCTCC 1621
Db 360 SerSerGlyAspValPheValIleProAlaGlyHisProValAlaValLysAlaSerSer 379
Qy 1622 GAATCCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1681
Db 380 AsnLeuAspLeuLeuGlyPheGlyIleAsnAlaGluAsnGlnArgAsnPheLeuAla 399
Qy 1682 GGTGATAGACATGTGATAGACCATGACATCAAAACCCAGAGGAATCTCACTTTGTGAGT 1741
Db 400 GlyAspGluAspAsnValIleSerGlnIleGlnArgProValLysGluLeuAlaPhePro 419
Qy 1742 GGGTCGGGTGAACAAGTTGAGAAGCTCATCAAAACCCAGAGGAATCTCACTTTGTGAGT 1801
Db 420 GlySerAlaGlnGluValAspArgIleLeuGluAsnGlnLysGlnSerHisPheAlaAsp 439
Qy 1802 GCTGCTCTCAATCTCAATCTCAATCTCAATCTCAATCTCAATCTCAATCTCAATCTCA 1861
Db 440 AlaGlnProGlnGlnArg-----GluArgGlySerArgGluThr 452
Qy 1862 GAGGATCAA 1870
Db 453 ArgAspArg 455

RESULT 6
S00567
vicilin precursor (clone pDUB9) - garden pea (fragment)
C;Species: Pisum sativum (garden pea)
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004
C;Accession: S00567
R;Watson, M.D.; Lambert, N.; Delaunev, A.; Yarwood, J.N.; Crox, R.R.D.; Gatehouse, J.A.;
Biochem. J. 251, 857-864, 1988
A;Title: Isolation and expression of a pea vicilin cDNA in the yeast Saccharomyces cerevisiae
A;Reference number: S00567; MUID:88326226; PMID:3046604
A;Molecule type: mRNA
A;Accession: S00567
A;Residues: 1-433 <WAT>
A;Cross-references: UNIPROT:P13918; EMBL:Y00722; NID:q20915; PIDN:CAA68708.1; PID:g75824
C;Superfamily: glycinin
F;1-2/Domain: signal sequence (fragment) #status predicted <SIG>
F;3-433/Product: vicilin #status predicted <MAT>

Alignment Scores:
Pred. No.: 2 5e-74 Length: 433
Score: 1174.50 Matches: 221
Percent Similarity: 72.18% Conservative: 93
Best Local Similarity: 50.80% Mismatches: 108
Query Match: 32.97% Indels: 13
DB: 2 Gaps: 4

US-10-728-051-1 (1-2032) x S00567 (1-433)

Qy 566 AACCTTTCTACTTCCTCGTCAAGCGGTTTGTAGCACCGCTACCGGACCAAAACCGGTAGG 625
Db 8 AsnProPheIlePheLysSerAsnLysPheGlnThrLeuPheGluAsnGluAsnGlyHis 27
Qy 626 ATCCGGTCTCTGACAGGTTTGACCAAGGTCAAGGAGGTTTCAGATCTCCAGATCAC 685
Db 28 IleArgLeuLeuGlnLysPheAspGlnArgSerLysIlePheGluAsnLeuGlnAsnTyr 47
Qy 686 CGTATTGTGACATCGAGGCGCAACCTCAACACTCTTCTTCTCCAGACGCTGATGCT 745
Db 48 ArgLeuLeuGluTyrLysSerLysProHisThrIlePheLeuProGlnHisThrAspAla 67
```

```
Qy 746 GATAACATCTTGTATTTCAGCAAGGCAAGCCACCGTGACCGTAGCAAAATGGCAATAAC 805
Db 68 AspTyrIleLeuValLeuSerGlyLysAlaIleLeuThrValLeuLysProAspAsp 87
Qy 806 AGAAGAGCTTTAATCTTGACGAGGCGCATCTCAGATCCATCCGCTTCATTTCC 865
Db 88 ArgAsnSerPheAsnLeuGluArgGlyAspThrIleLysLeuProAlaGlyThrIleAla 107
Qy 866 TACATCTTGAACCGCCATCAACACCAAGCACTCAGATAGCTAAAATCTCCATGCCCGTT 925
Db 108 TyrLeuValAsnArgAspAsnGluLeuArgValLeuAspLeuAlaIleProVal 127
Qy 926 AACACACCGCGCCAGTTTGAGGATTTCTCCCGGAGCAGCGAGCAACCAATCATCTAC 985
Db 128 AsnArgProGlyGlnLeuGlnSerPheLeuLeuSerGlyAsnGlnAsnGlnAsnTyr 147
Qy 986 TTCACGGCTTTCAGCAGGATACGTTGGAGGCGCTTCAATCGGGAATTCATGAGATA 1045
Db 148 LeuSerGlyPheSerLysAsnIleLeuGluAlaSerPheAsnThrAspTyrGluGluIle 167
Qy 1046 CGAGGGTGTGTAGAGAGAAATGCAGAGGTGAGCAAGAGAGAGAGGCGCAGAGCGGA 1105
Db 168 GluLysValLeuLeuGluHisGluLysGluThrGlnHisArgArgSerLeuLys--- 186
Qy 1106 TGGAGTACTCGGAGTAGTGAGAACAAATGAGGAGTAGTGTCAAGTGTCAAAGGAGCAC 1165
Db 187 ---AspLysArgGlnGlnSerGlnGluGluAsnValIleValLysLeuSerArgGlyGln 205
Qy 1166 GTTGAGAACTTACTAAGCAGCTAAATCCGTCTCAAGAAAGGCTCCGAAGAAGGCGGA 1225
Db 206 IleGluGluLeuSerLysAsnAlaLysSerThrSerLysLysSerValSerSerGluSer 225
Qy 1226 GATATCAACCAACCAATCAACTTGAGAGAGGCGGAGCCCATCTTTTCAACACTTTGGG 1285
Db 226 Glu-----ProPheAsnLeuArgSerArgGlyProIleTyrSerAsnGluPheGly 242
Qy 1286 AAGTTATTGAGTGAAGCAGACAGAAAGAACCCAGCTTCAGGACCTTGACATGATG 1345
Db 243 LysPhePheGluIleThrProGlu---LysAsnProGlnLeuGlnAspLeuAspIlePhe 261
Qy 1346 CTCACCTGTGTAGAGATCAAGAGGAGCTTGTATGCTCCACACTTCACTCAAAAGGCC 1405
Db 262 ValAsnSerValGluLysGluSerLeuLeuLeuProHisTyrAsnSerArgAla 281
Qy 1406 ATGCTTATCGTCGTCTCAACAAAGAACTGGAACCTTCAACTCGTGGCTGTAAAGAAA 1465
Db 282 IleValIleValThrValAsnGluGlyLysGlyAspPheGluLeuValGlyGlnArgAsn 301
Qy 1466 GAGCAACAACAGAGGAGCGCGGGAAGAGAGGAGACGAAAGAGAGAGAGGAGGGA 1525
Db 302 GluAsnGlnGlnGluGlnArgLysGluAspGluGluGluGluGluGluGlu 321
Qy 1526 AGTAAACAGAGAGGTCGTAGGTACACAGCGAGGTTCAGAGGAGGCTGTTCATCATG 1585
Db 322 IleAsnLysGlnValGlnAsnTyrLysAlaLysLeuSerSerGlyAspValPheValIle 341
Qy 1586 CCAGAGCTCATCCAGTAGCCATCAACGCTTCCTCCGAACTCCATCTGCTTGGCTTCGGT 1645
Db 342 ProAlaGlyHisProValAlaLeuLysAlaSerSerAsnLeuAspLeuLeuPheGly 361
Qy 1646 ATCAACGCTGAAAAACCAACCAAGATCTTCTTTCAGGAGGTATAGGCAATGTGATAGAC 1705
Db 362 IleAsnAlaGluAsnAsnGlnArgAsnPheLeuAlaGlyAspGluAspAsnValIleSer 381
Qy 1706 CAGATAGAGAGCAACGAGGAGGATTTAGCATTCCTCGGTCCGGTGAACAAGTTGAGAAG 1765
Db 382 GlnValGlnArgProValLysGluLeuAlaPheProGlySerAlaGlnGluValAspArg 401
Qy 1766 CTCATCAAAAACCAAGAGGAAATCTCACTTTGTGAGTCTGCTCTCAATCTCAATCTCAA 1825
Db 402 IleLeuGluAsnGlnLysGlnSerHisPheAlaAspAlaGlnProGlnGlnArg----- 419
```

QY 1826 TCTCCGTCGTCCTCTGAGAAGAGTCTCTCTGAGAAGAGCATCAA 1870
Db 420 -----GluArgGlySerArgGluThrArgAspArg 429
RESULT 7
FMSYCB
beta-conglycinin beta chain - soybean
C:Species: Glycine max (soybean)
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 09-Jul-2004
C:Accession: JQ0969
R:Harada, J.J.; Barker, S.J.; Goldberg, R.B.
Plant Cell 1, 415-425, 1989
A:Title: Soybean beta-conglycinin genes are clustered in several DNA regions and are reg
A:Reference number: JQ0969; MUID:93005638; PMID:2562562
A:Accession: JQ0969
A:Molecule type: DNA
A:Residues: 1-439 <HAR>
A:Cross-references: UNIPROT:P25974; GB:S44893; NID:G256426; PID:G256427
C:Comment: This protein accumulates during seed development and is hydrolyzed after germ
C:Genetics:
A:Gene: CG-4
A:Introns: 101/1; 159/3; 186/3; 281/3; 375/1
C:Superfamily: glycinin
C:Keywords: seed; storage protein
Alignment Scores:
Pred. No.: 4,95e-73 Length: 439
Score: 1156.00 Matches: 237
Percent Similarity: 69.98% Conservative: 87
Best Local Similarity: 51.19% Mismatches: 85
Query Match: 32.45% Indels: 54
DB: 1 Gaps: 10
US-10-728-051-1 (1-2032) x FMSYCB (1-439)
QY 542 GTGAGGGAAGAAACATCTCGSAACACCCCTTTCTACTTC---CCGTCAAGCGCGTTTACG 598
Db 26 ValArgGluAsp-----GluAsnAsnProPheTyrPheArgSerSerAsnSerPheGln 43
QY 599 ACCCGTACGGGAACCAAAACGGTAGATCCGGTCTCTGCAGAGGTTTGACCAAGTCA 658
Db 44 ThrLeuPheGluAsnGlnAsnValArgLleArgLeuLeuGlnArgPheAsnLysArgSer 63
QY 659 AGGCAGTTTCCAGATCTCCAGATCACCGTATTGTGCAGATCGAGGCCAACCTAACACT 718
Db 64 ProGlnLeuGluAsnLeuArgAspTyrArgLleValGlnPheGlnSerLysProAsnThr 83
QY 719 CTTGTTCTTCCAGACGCGTGTGATTAACATCTGTTATCTCCAGCAAGCGCAAGCC 778
Db 84 IleLeuLeuProHisAlaAspAlaAspPheLeuLeuPheValLeuSerGlyArgAla 103
QY 779 ACCGTGACCGTAGCAATTCGCAATAACAGAAAGAGCTTTAATCTTGCAGGGCGCATGCA 838
Db 104 IleLeuThrLeuValAsnAsnAspArgAspSerTyrAsnLeuHisProGlyAspAla 123
QY 839 CTCAGATCCCATCCGGTTTCATTCTCTACATCTTGAACCGCCATGACACCAAGCACTC 898
Db 124 GlnArgLleProAlaGlyThrTyrTyrLeuValAsnProHisAspHisGlnAsnLeu 143
QY 899 AGAGTAGCTAAATCTCCATCCCGTGTAAACACCCGCCAGTGTGAGGATTTCTCCCG 958
Db 144 LysIleLeuLeuAlaIleProValAsnLysProGlyArgTyrAspAspPhePheLeu 163
QY 959 GCGAGCAGCGAGACCAATCATCTTACTTGAGGGCTTCAGCAGGAATACGTTGAGGCC 1018
Db 164 SerSerThrGlnAlaGlnGlnSerTyrLeuGlnGlyPheSerHisAsnIleLeuGluThr 183
QY 1019 GCCTTCAATCGGAATCAATGAGATACGGAGGCTGTGTTAGAGAGAAATGCAGAGAGT 1078
Db 184 SerPheHisSerGluPheGluGluIleAsnArgValLeuPhe-----Gly 198
QY 1079 GAGCAAGAGGAGAGGCGCAGCGATGGAGTAGTACTCGGAGTAGTGAGAAACAATGAAGGA 1138
Db 1138 -----Gly 198


```
QY 854 GGTTCATTCTCATCTTGAACCGCCATGACAAACAGACCTCAGAGTAGCTAAATC 913
Db 130 GlyThrIleGlyTyrLeuValAsnArgAspGluGluAspLeuArgValLeuAspLeu 149
QY 914 TCCATGCCGTTAAACACCCGGCAGTTTGAGATTTCTCCGGCGAGCAGCCGAGAC 973
Db 150 ValIleProValAsnArgProGlyGluProGlnSerPheLeuLeuSerGlyAsnGln 169
QY 974 CAATCATCTACTTGCAGGGCTTGCAGAGGATACGTTGGAGCGCGCTTCAATGCCGA 1033
Db 170 GlnProSerIleLeuSerGlyPheSerLysAsnIleLeuGluAlaSerPheAsnThrAsp 189
QY 1034 TTCAATGAGATACCGGAGGCTCTGTAGAGAGATGTCAGGAGGTGAGCAAGAGGAGA 1093
Db 190 TyrLysGluIleGluLeuValLeuGluGluHisGlyLysGluLysTyrHisArgArg 209
QY 1094 GGGCAGAGCGCATGAGTACTCGAGTAGTGTAGAACAAATGAGAGGTGATAGTCAAGTG 1153
Db 210 GlyLeuLysAspArgA-gGlnArgGlyGlnGluAsn-----ValIleValLysIle 227
QY 1154 TCAAAGAGCAGCGTTGAAGAACTTACTAAGCACCTAAATCCGTCCTCAAAGAAAGGCTCC 1213
Db 228 SerArgLysGlnIleGluGluLeuAsnLysAsnLysSerSerLysLysSerThr 247
QY 1214 GAAGAAGAGGAGATATACCAACCCCAATCAACTTGAGAGAAGCGAGCCGATCTTTCT 1273
Db 248 SerSerGluSerGlu-----ProPheAsnLeuLysSerArgGluProIleTyrSer 264
QY 1274 AACAACTTGGAGAGTATTGAGGTGAAGCCACAGCAAGAAACCCCGAGCTTCAGGAC 1333
Db 265 AsnLysPheGlyLysPhePheGluIleThrPro---LysArgAsnProGlnLeuGlnAsp 283
QY 1334 CTGGACATGATGCTCACCTGTGTAGATCAAAAGAGAGCTTTGATGCTCCACACTTC 1393
Db 284 LeuAsnIlePheValAsnTyrValGluIleAsnGluGlySerLeuLeuLeuProHisTyr 303
QY 1394 AACTCAAGGCGCATGTTATGCTGCTGCTCAACAAGAGAACTGGAACCTTGAACCTCGTG 1453
Db 304 AsnSerArgAlaIleValIleValThrValAsnGluGlyLysGlyAspPheGluLeuVal 323
QY 1454 GCTGTAGAAAAGAGCAACACAGAGGGGAGCGGGAAGCAAGAGGAGGAGCAAGCA 1513
Db 324 GlyLysArgAsnGluAsnGlnGlnGlyLeuArgGluGluTyrAspGluGluLysGluGln 343
QY 1514 GAGAGAGGAGAACTAACAGAGAGTCCGTAGGTACACAGCGAGGTGTAAGGAGCGAT 1573
Db 344 GlyGluGluIleArgLysGlnValGlnAsnTyrLysAlaLysLeuSerProGlyAsp 363
QY 1574 GTGTTTCATGTCAGCAGCTCATCCAGTAGCCATCAACGCTTCCCGAATCCCATCTG 1633
Db 364 ValLeuValIleProAlaGlyTyrProValAlaIleLysAlaSerSerAsnLeuLeu 383
QY 1634 CTTGGCTTCGGTATCAACGCTGAAACCAACACAGAACTCTCTTCAGTGTATAGGAC 1693
Db 384 ValGlyPheGlyIleAsnAlaGluAsnAsnGlnArgTyrPheLeuAlaGlyGluGluAsp 403
QY 1694 AATGTGATGACGAGATAGAGAGCAAGCAAGATTTAGCATTCCTCGGTGGGTGAA 1753
Db 404 AsnValIleSerGlnIleHisLysProValLysGluLeuAlaPheProGlySerAlaGln 423
QY 1754 CAAGTTGAGAGCTCATCAAAACACAGAGGATCTCACTTTGTAGTCTGCTCTCAA 1813
Db 424 GluValAspThrLeuLeuGluAsnGlnLysGlnSerHisPheAlaAsnAlaGlnProGln 443
QY 1814 TCTCAATCTCAATCTCCGTCGTCTCCTGAGAAAGAGTCTCCTGAGAAAGAGATCAAG 1873
Db 444 -----GluArgGlu 446
QY 1874 GAGAAAACCAAGGAGGAGGTTCATCTCTTCAATTTTGAAGCTTTT 1924
Db 447 ArgGlySerGlnGluIleLysAspHisLeuTyrSerIleLeuGlySerPhe 463
```

RESULT 10

```
S06309
vicilin precursor (clone VFVic1) - tick bean
N;Alternate names: 7S seed storage protein
C;Species: Vicia faba var. minor (tick bean)
C;Date: 31-Mar-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004
C;Accession: S06309; S06456
R;Weschke, W.; Baumelein, H.; Wobus, U.
Nucleic Acids Res. 15, 10085, 1987
A;Title: Nucleotide sequence of a field bean (Vicia faba L. var. minor) vicilin gene.
A;Reference number: S06309; MUID:88096511; PMID:3697075
A;Accession: S06309
A;Molecule type: DNA
A;Residues: 1-463 <WES>
A;Cross-references: UNIPROT:P08438; EMBL:Y00506; NID:9829146; PID:CAA68559.1; PID:g2205
A;Note: the authors translated the codon TCT for residue 296 as Cys
R;Weschke, W.; Bassuener, R.; van Hal, N.; Czihal, A.; Baumelein, H.; Wobus, U.
Biochem. Physiol. 183, 233-242, 1988
A;Title: The structure of a Vicia faba vicilin gene.
A;Reference number: S06456
A;Accession: S06456
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-463 <WES2>
A;Note: 257-Lys and 443-His were also found
C;Genetics:
A;Introns: 102/1; 160/3; 187/3; 295/3; 400/1
C;Superfamily: glycinin
C;Keywords: seed; storage protein
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-463/Product: vicilin #status predicted <MAT>

Alignment Scores:
Pred. No.: 1,06e-69 Length: 463
Score: 1108.50 Matches: 218
Percent Similarity: 68.93% Conservative: 97
Best Local Similarity: 47.70% Mismatches: 119
Query Match: 31.12% Indels: 23
DB: 2 Gaps: 4

US-10-728-051-1 (1-2032) x S06309 (1-463)
```

```
QY 554 ACATCTCGAACAACCTTTCTTCTCCGTCAGCGGTTTAGCACCGCTACCGGAAC 613
Db 30 SerAspGlnAsnProPheValPheGluSerAsnArgPheGlnThrLeuPheGluAsn 49
QY 614 CAAAACGCTAGATCGGGTCTCGAGAGTTTGACAAAGTCAAGGAGTTTCAGAAAT 673
Db 50 GluAsnGlyHisIleArgLeuLeuGlnLysPheAspGlnHisSerLysLeuLeuGluAsn 69
QY 674 CTCAGAATCACCGTATTGTGCAGATCGAGGCCAAACCTAACACTCTTGTCTTCCCAAG 733
Db 70 LeuGlnAsnTyrArgLeuLeuGluTyrLysSerLysProHisThrIlePheLeuProGln 89
QY 734 CACGCTGATGTGATPACATCTTGTATTATCCAGCAAGGCCAAGCCACCGTACCGTAGCA 793
Db 90 GlnThrAspAlaAspPheIleLeuValLeuSerGlyLysAlaIleLeuThrValLeu 109
QY 794 AATGCAATAACAGAAAGCTTTAATCTTCAGGAGGCCATGACTCAGATCCCATCC 853
Db 110 LeuProAsnAspArgAsnSerPheSerLeuGluArgGlyAspThrIleLysLeuProAla 129
QY 854 GGTTCATTCTCATCTTGAACCGCCATGACAAACCAACCTCAGAGTAGTAGTAAAAATC 913
Db 130 GlyThrIleGlyTyrLeuValAsnArgAspAspGluGluAspLeuArgValLeuAspLeu 149
QY 914 TCCATGCCGTTAAACACCCGGCAGTTTGGAGATTTCTCCCGGCGAGCGGAGAC 973
Db 150 ValIleProValAsnArgProGlyGluProGlnSerPheLeuLeuSerGlyAsnGlnAsn 169
QY 974 CAATCATCTACTTGCAGGCTTGCAGAGGATACGTTGGAGGCGGCTTCAATCGCGAA 1033
Db 170 GlnProSerIleLeuSerGlyPheSerLysAsnIleLeuGluAlaSerPheAsnThrAsp 189
```

```
QY 1034 TTCATAGATACGAGGGTCTGTTTAAAGAGAAATCAGAGAGTGTAGCAGAGCAGAGA 1093
Db 190 TyrlsGluilleGluLysValLeuLeuGluGluHisGlyLysGluLysTyrHisArgArg 209
QY 1094 GGCAGAGGCGATGAGTACTCGAGTAGTGAGAACAAATGAGGAGGTATGATCAAGTG 1153
Db 210 GlyLeuLysAspArgArgGlnArgGlyGlnGluGluAsn-----ValIleValLysIle 227
QY 1154 TCAAGAGGACGCGTTGAAGAACTTACTAAGCACGCTAAATCCGCTCAAAAGAAAGGCTCC 1213
Db 228 SerArgLysGlnIleGluLeuLeuAsnLysAsnAlaLysSerSerLysLysSerThr 247
QY 1214 GAGAGAGGAGATATCAACAACCAATCAACTTGAGAGAGCGCCGATCTTTCT 1273
Db 248 SerSerGluSerGlu-----ProPheAsnLeuArgSerArgGluProIleTyrSer 264
QY 1274 AACAACTTGGGAGTATTGAGGTGAGCCAGCAAGAACCCCGAGCTTCAGGAC 1333
Db 265 AsnLysPheGlyLysPhePheGluIleThrPro---LysArgAsnProGlnLeuGlnAsp 283
QY 1334 CTGACATGATGCTCACCTGTGTAGATCAAGAAGGAGCTTTGATGCTCCACACTTC 1393
Db 284 LeuAsnIlePheValAsnTyrValGluIleAsnGluGlySerLeuLeuLeuProHisTyr 303
QY 1394 AACTCAAAGCCATGTTGCTGCTGCTCAACAAGAACTTGAACCTTGAACCTGCTG 1453
Db 304 AsnSerArgAlaIleValIleValThrValAsnGluGlyLysGlyAspPheGluLeuVal 323
QY 1454 GCTGTAGAAGAGCAACACAGAGGGGCGCGGGAAGAGAGGAGCAGAGCGAA 1513
Db 324 GlyLysArgAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 343
QY 1514 GAAGAGAGGAGGAGTAAACAGAGAGTGTAGGTACACAGAGGTGAAGGAGGAGAT 1573
Db 344 GlyGluGluGluLeuArgLysGlnValGlnAsnTyrLysAlaLysLeuSerProGlyAsp 363
QY 1574 GTGTTATCATGACGAGCTCATCCAGTAGCCATCAACGCTTCTCCGAACTCCATCTG 1633
Db 364 ValLeuValIleProAlaGlyTyrProValAlaIleLysAlaSerSerAsnLeuAsnLeu 383
QY 1634 CTGTCCTCGGTATCAAGCTGAACAAACACAGATCTCTCTGAGGTGATAGGAC 1693
Db 384 ValGlyPheGlyLysAsnAlaGluAsnAsnGlnArgTyrPheLeuAlaGlyGluGluAsp 403
QY 1694 AATGTGATAGACAGATAGAGAGCAAGCGAAGATTAGATTCCCTCGGTGCGGTGAA 1753
Db 404 AsnValIleSerGlnIleHisLysProValLysGluLeuAlaPheProGlySerAlaGln 423
QY 1754 CAAGTTGAGAGGTATCAAAAACCAAGAGAAATCTCATTGTGAGTGTCTGCTCTCAA 1813
Db 424 GluValAspThrLeuLeuGluAsnGlnLysGlnSerHisPheAlaAsnAlaGlnProArg 443
QY 1814 TCTCAATCTCAATCTCGTGTCTCTGAGAGAGAGTCTCTGAGAAAGAGATCAAG 1873
Db 444 -----GluArgGlu 446
QY 1874 GAGGAAACCAAGAGGAGGAGGTCCTCTTCAATTTTGAAGGCTTTT 1924
Db 447 ArgGlySerGlnGluIleLysAspHisLeuTyrSerIleLeuGlySerPhe 463

RESULT 11
FWPMVB
vicilin B precursor - garden pea (fragment)
C:Species: Pisum sativum (garden pea)
C:Date: 14-Nov-1983 #sequence_revision 14-Nov-1983 #text_change 09-Jul-2004
C:Accession: A03344
R:Lycett, G.W.; Delauney, A.J.; Gatehouse, J.A.; Gilroy, J.; Croy, R.R.D.; Boulter, D.
Nucleic Acids Res. 11, 2367-2380, 1983
A:Title: The vicilin gene family of pea (Pisum sativum L.): a complete cDNA coding sequence
A:Reference number: A93462; MUID:83220791; PMID:6687941
A:Accession: A03344
A:Molecule type: mRNA
```

A;Residues: 1-410 <LYC>

A;Cross-references: UNIPROT:P02854

A;Experimental source: cv. Feltham First, clones pDBU7 and pDBU4

A;Note: parts of this sequence, including the amino end of the mature protein, were determined from a cDNA library

C;Comment: The gene that codes for this protein is part of a multigene family coding for

C;Superfamily: glycinin

C;Keywords: glycoprotein

F;1-15/Domain: signal sequence #status predicted <SIG>

F;16-410/Product: vicilin type B (fragment) #status predicted <MAT>

F;321-322/Cleavage site: Asn-Asp (unidentified proteinase) #status experimental

F;359/Binding site: carbohydrate (Asn) (covalent) #status experimental

Alignment Scores:

Pred. No.: 7,05e-67 Length: 410

Score: 1068.00 Matches: 209

Percent Similarity: 71.64% Conservative: 84

Best Local Similarity: 51.10% Mismatches: 104

Query Match: 29.98% Indels: 12

DB: 1 Gaps: 5

US-10-728-051-1 (1-2032) x FWPMB (1-410)

QY 542 GTGAGGAGAGAAACATCTCGAACAACCTTTCTACTTCCGCTCAAGCGGTTTAGCACC 601

Db 13 ValSerSerArgSerAspGlnGluAsnProPheIlePheLysSerAsnArgPheGlnThr 32

QY 602 CGCTACGGGAACCAAAACGGTAGGATCCGGGTCTCGAGAGGTTTGACCAAGGTCAAGG 661

Db 33 LeuTyrGluAsnGluAsnGlyHisIleArgLeuLeuGlnLysPheAspLysArgSerLys 52

QY 662 CAGTTTCAGAACTCCAGAAATCACCGTATGTCAGATCGAGCCGCAACCTTAACACTCTT 721

Db 53 IlePheGluAsnLeuGlnAsnTyrArgLeuLeuGlyTyrLysSerLysProHisThrLeu 72

QY 722 GTTCTTCCCAAGCACGCTGATGCTGATACATCTCTGTTATCCAGCAAGGCAAGCACC 781

Db 73 PheLeuProGlnTyrThrAspAlaAspPheIleLeuValLeuSerGlyLysAlaThr 92

QY 782 GTACCGGTAGCAATGCAATAACAGAAAGGCTTTAATCTTCAGAGGCGCATGCACTC 841

Db 93 LeuThrValLeuLysSerAsnAspArgAsnSerPheAsnLeuGluArgGlyAspAlaIle 112

QY 842 AGAATCCCATCCGGTTTCATTTCTACATCTTGACCGCCATGACAAACCAAGCTCAGA 901

Db 113 LysLeuProAlaGlySerIleAlaTyrPheAlaAsnArgAspAsnGluGluProArg 132

QY 902 GTAGCTAAATCTCCATGCGCTTAACACACCGCGCAGTTTCAGGATTTCTCCCGCG 961

Db 133 ValLeuAspLeuAlaIleProValAsnLysProGlyGlnLeuGlnSerPheLeuLeuSer 152

QY 962 AGCAGCGAGACCAATCATCTTACTTGCAGGCTTCAGCAGGAATACGTTGGAGCGCGCC 1021

Db 153 GlyThrGlnAsnGlnLysSerSerLeuSerGlyPheSerLysAsnIleLeuGluAlaAla 172

QY 1022 TTCATTCGGAATTCATAGATACGAGGCTGTTTGAAGAGAAATGCGAGGTTGAG 1081

Db 173 PheAsnThrAsnTyrGluGluIleGluLysValLeuLeuGlnGln-----Glu 189

QY 1082 CAGAGGAGAGAGCGGAGCGCATGGAGTACTCGGAGTAGTAGG---AACAAATGAGGA 1138

Db 190 GlnGluProGlnHisArgSerSerLeuLysAspArgGlnGluIleAsnGluGluAsn 209

QY 1139 GTGATAGTCAAAGTCAAGGAGACGCTTGAAGAATCTACTAAGCACGCTAAATCCGTC 1198

Db 210 ValIleValLysValSerArgAspGlnIleGluLeuSerLysAsnAlaLysSerSer 229

QY 1199 TCAAGAAAGGCTCCGAAGAGAGGAGATATCACCAACCCCAATCAACTTGAAGAGGC 1258

Db 230 SerLysLysSerValSerSerGlu-----SerGlyProPheAsnLeuArgSerArg 246

QY 1259 GAGCCCGCATCTTCTAACAACCTTTGGAGAGTATTTCAGGTGAAGCCAGACAGAAAC 1318

Db 247 AsnProIleTyrSerAsnLysPheGlyLysPhePheGluIleThrProGlu---LysAsn 265


```
QY 1208 GGCTCCGAAGAGGAGATATATACCAACCCCAATCACTTGTAGAGAGGCGGAGCCCGAT 1267
Db 241 ThrLeuSerSerGln-----AspLysProPheAsnLeuArgSerArgAspProIle 257
QY 1268 CTTTCTAACAACTTTGGGAAGTTATTGAGGTGAAGCCACAGCAAGAGAGACCCCGAGCTT 1327
Db 258 TyrSerAsnAsnTyrGlyLysLeuTyrGluIleThrProGlu---LysAsnSerGlnLeu 276
QY 1328 CAGGACCTGCACATGATGCTCACTGTGTAGAGATCAAAAGAGAGCTTTGATGCTCCCA 1387
Db 277 ArgAspLeuAspIleLeuLeuAsnCysLeuGlnMetAsnGluGlyAlaLeuPheValPro 296
QY 1388 CACTTCAACTCAAGGCCATGCTTATCGTGTCTCAACAAGGAAGAACTGGAACCTTGAA 1447
Db 297 HisTyrAsnSerArgAlaThrValIleLeuValAlaAsnGluGlyArgAlaGluValGlu 316
QY 1448 CTGCTGGCTGTAGAAAAGAGCAACAACAGAGGGGCGCGGGAAGAAGAGGAGGACGAA 1507
Db 317 LeuValGlyLeuGluGlnGlnGln----- 326
QY 1508 GACGAAGAAGAGGAGGA---AGTAACAGAGAGTGCCTAGGTACACAGCGAGTTGAAG 1564
Db 327 -----GlyLeuGluSerMetGlnLeuArgTyrAlaAlaThrLeuSer 341
QY 1565 GAAGGCATGTTTCATCATGCCAGCAGCTCATCCAGTACCATCAACGCTTCTCCGAA 1624
Db 342 GluGlyAspIleIleValIleProSerSerPheProValAlaLeuLysAlaAlaSerAsp 361
QY 1625 CTCATCTGCTTGGCTTCGGTATCAACGCTGAAACCAACACACAGAACTTCTTTCGAGGT 1684
Db 362 LeuAsnMetValGlyIleGlyValAsnAlaGluAsnGluArgAsnPheLeuAlaGly 381
QY 1685 GATAAGGCAATGTGTAGACCATGATAGAGCAAGCGAAGGATTTAGCATTTCCCTGGG 1744
Db 382 HisLysGluAsnValIleArgGlnIleProArgGlnValSerAspLeuThrPheProGly 401
QY 1745 TCGGTGAACAAGTTGAGAGCTCATCAAAACAGCAAGAAATCTCACTTTGTGAGTGT 1804
Db 402 SerGlyGluGluValGluGluLeuLeuGluAsnGlnLysGluSerTyrPheValAspGly 421
QY 1805 CGTCCTCAA 1813
Db 422 GlnProArg 424

RESULT 13
S00281
canavalin - sword bean
C:Species: Canavalia gladiata (sword bean)
C>Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 09-Jul-2004
C:Accession: S00281; S04598
R:Yamauchi, D.; Nakamura, K.; Asahi, T.; Minamikawa, T.
Eur. J. Biochem. 170, 515-520, 1988
A:Title: cDNAs for canavalin and concanavalin A from Canavalia gladiata seeds. Nucleotide
s.
A:Reference number: S00281; MUID:88111636; PMID:3338449
A:Accession: S00281
A:Molecule type: mRNA
A:Residues: 1-445 <YAM>
A:Cross-references: UNIPROT:P10562; EMBL:X06733; NID:g18003; PIDN:CAA29910.1; PID:g18004
R:Takei, Y.; Yamauchi, D.; Minamikawa, T.
Nucleic Acids Res. 17, 4381, 1989
A:Title: Nucleotide sequence of the canavalin gene from Canavalia gladiata seeds.
A:Reference number: S04598; MUID:89296493; PMID:2740227
A:Accession: S04598
A:Molecule type: DNA
A:Residues: 1-160, 'K', 162-445 <TAK>
A:Cross-references: EMBL:X15076; NID:g18006; PIDN:CAA33172.1; PID:g18007
C:Genetics:
A:Introns: 115/1; 173/3; 200/3; 290/3; 381/1
C:Superfamily: glycinin
C:Keywords: seed

Alignment Scores:
```

```
Pred. No.: 2,82e-59 Length: 445
Score: 959.50 Matches: 197
Percent Similarity: 54.03% Conservative: 118
Best Local Similarity: 33.79% Mismatches: 101
Query Match: 26.94% Indels: 167
DB: 2 Gaps: 8

US-10-728-051-1 (1-2032) x S00281 (1-445)
QY 68 CCACTCATGCTGTGTAGGATCCTTGTCTCGCTTCACTTCTTGTGCAAGCGATGCCAAG 127
Db 8 ProLeuTrpLeuLeuLeuGlyValValLeuLeuAlaSerValSerAlaSerPheAlaHis 27
QY 128 TCATCACTTTACAGAAAGAAACAGAGAACCCCTGCGCCACAGAGTGCCTCCAGAGTTGT 187
Db 28 Ser----- 28
QY 188 CAACAGNACCGGATGACTTGAAGCAAAAGCATGCGAGTCTCGCTGCACCAAGCTCGAG 247
Db 28 ----- 28
QY 248 TATGATCCTCGTGTGTCTATGATCCTCGAGGACACACTGCCACCAACCAACGTTCC 307
Db 29 -----GlyHisSerGly----- 32
QY 308 CTTCAGGGAGCGGACACAGTGGCCGCAACCCGAGACTACGATGATGACGCGCTCAA 367
Db 33 -----GlyGluAlaGluAspGluSerGluGlu 41
QY 368 CCCGAAGAGAGAGGAGGCCGATGGGACAGCTGGAGCGGAGGCGTGAAGAGAA 427
Db 42 SerArg----- 43
QY 428 GAAGACTGGAGACAAACAGAGAAAGATTGGAGGCGCAAGTCAATCAGACGACCGGAAA 487
Db 43 ----- 43
QY 488 ATAAGCCCGAAGAAAGAGAGAGAAACAAGAGTGGGGAAACACAGGTAGCCATGTGAGG 547
Db 43 ----- 43
QY 548 GAAGAAACATCTCGGAACACCCCTTTCTACTTCCGTCAGGCGGTTTAGCACCCGCTAC 607
Db 44 -----AlaGlnAsnAsnProTyrLeuPheArgSerAsnLysPheLeuThrLeuPhe 60
QY 608 GGGAAACAAACCGTAGGATCCGGCTCTGACAGAGTTTGCACCAAGGTCAAGGCAGTTT 667
Db 61 LysAsnGlnHisGlySerLeuArgLeuLeuGlnArgPheAsnGluAspThrGluLysLeu 80
QY 668 CAGAACTCCAGAAATCACCGTATTGTGCAGATCGAGGCCAAACCTAACACTTCTTCTT 727
Db 81 GluAsnLeuArgAspTyrArgValLeuGluTyrCysSerLysProAsnThrLeuLeuLeu 100
QY 728 CCCAAGCAGCTCATCTGATACATCTTGTATCCAGCAAGGCGCAAGCCCGTCAACC 787
Db 101 ProHisHisSerAspSerAspLeuLeuValLeuGluGlyGlnAlaIleLeuVal 120
QY 788 GTAGCAATGGCAATAACAGAAAGAGCTTTAACTTTGACGAGGCGCATGCACTCAGAAATC 847
Db 121 LeuValAsnProAspGlyArgAspThrTyrLysLeuAspGlnGlyAspAlaIleLysIle 140
QY 848 CCATCCGTTTCAATTTCTCATCTTTGAACCGCCATGACCAACAGAACTCCAGAGTAGCT 907
Db 141 GlnAlaGlyThrProPheTyrLeuIleAsnProAspAsnAsnGlnAsnLeuArgIleLeu 160
QY 908 AAATCTCCATGCCCGTTAAACACACCGCGCAGCTTTGAGGATTTCTCCGCGGAGCAGC 967
Db 161 AsnPheAlaIleThrPheArgArgProGlyThrValGluAspPheLeuSerSerThr 180
QY 968 CGAGACCAATCATCTTACTTTCAGGCGCTTCAGCAGGAATACGTTGGAGGCGCGCTTCAAT 1027
Db 181 LysArgLeuProSerTyrLeuSerAlaPheSerLysAsnPheLeuGluAlaSerTyrAsp 200
```



```
QY 1028 GCGGAATTCATAGATACGAGGAGGTGCTGTAGAGAGAAATGCAGGAGGTGAGCAAGAG 1087
Db 1028 GCGGAATTCATAGATACGAGGAGGTGCTGTAGAGAGAAATGCAGGAGGTGAGCAAGAG 1087
QY 201 SerProTyrAspGluIleGluGlnThrLeuLeuGln----- 212
Db 201 SerProTyrAspGluIleGluGlnThrLeuLeuGln----- 212
QY 1088 GAGAGAGGCGAGAGCGGATGAGTACTCGGAGTAGTAGAGAACATGAGAGGTGATAGTC 1147
Db 1088 GAGAGAGGCGAGAGCGGATGAGTACTCGGAGTAGTAGAGAACATGAGAGGTGATAGTC 1147
QY 213 -----GluGluGlnGluGlyValIleVal 220
Db 213 -----GluGluGlnGluGlyValIleVal 220
QY 1148 AAAGTGTCAAAGGAGCAGCTTGAAGAACTTAAAGCAGCGTAAATCCGTCTCAAGAAA 1207
Db 1148 AAAGTGTCAAAGGAGCAGCTTGAAGAACTTAAAGCAGCGTAAATCCGTCTCAAGAAA 1207
QY 221 LysMetProLysAspGlnIleGlnGluIleSerLysHisAlaGlnSerSerArgLys 240
Db 221 LysMetProLysAspGlnIleGlnGluIleSerLysHisAlaGlnSerSerArgLys 240
QY 1208 GGTCCGGAAGAGGAGGAGATATACCAACCCAACTTACAGAGAGGCGAGCCCGAT 1267
Db 1208 GGTCCGGAAGAGGAGGAGATATACCAACCCAACTTACAGAGAGGCGAGCCCGAT 1267
QY 241 ThrLeuSerSerGln-----AspLysProPheAsnLeuArgSerArgAspProIle 257
Db 241 ThrLeuSerSerGln-----AspLysProPheAsnLeuArgSerArgAspProIle 257
QY 1268 CTTTCTAACAACTTTGGGAAGTATTATGAGTGAAGCCAGCAGACAGAGAACCCCGACTT 1327
Db 1268 CTTTCTAACAACTTTGGGAAGTATTATGAGTGAAGCCAGCAGACAGAGAACCCCGACTT 1327
QY 258 TyrSerAsnAsnTyrGlyLysLeuTyrGluIleThrProGlu-----LysAsnSerGlnLeu 276
Db 258 TyrSerAsnAsnTyrGlyLysLeuTyrGluIleThrProGlu-----LysAsnSerGlnLeu 276
QY 1328 CAGGACCTGGACATGCTCACCTGTGTAGAGATCAAGAGAGGCTTTGATGCTCCCA 1387
Db 1328 CAGGACCTGGACATGCTCACCTGTGTAGAGATCAAGAGAGGCTTTGATGCTCCCA 1387
QY 277 ArgAspLeuAspIleLeuLeuAsnCysLeuGlnMetAsnGluGlyAlaLeuPheValPro 296
Db 277 ArgAspLeuAspIleLeuLeuAsnCysLeuGlnMetAsnGluGlyAlaLeuPheValPro 296
QY 1388 CACTTCAACTCAAAGGCCAGTGTATCGTCTGCTCAACAAAGAACTGGAAACCTTGAA 1447
Db 1388 CACTTCAACTCAAAGGCCAGTGTATCGTCTGCTCAACAAAGAACTGGAAACCTTGAA 1447
QY 297 HisTyrAsnSerArgAlaThrValIleLeuValAlaAsnGluGlyArgAlaGluValGlu 316
Db 297 HisTyrAsnSerArgAlaThrValIleLeuValAlaAsnGluGlyArgAlaGluValGlu 316
QY 1448 CTCGTGGCTGTAAAGAAAGAGCAACACAGAGGCGCGCGGGAAGAGAGGAGCGAA 1507
Db 1448 CTCGTGGCTGTAAAGAAAGAGCAACACAGAGGCGCGCGGGAAGAGAGGAGCGAA 1507
QY 317 LeuValGlyLeuGluGlnGlnGln----- 326
Db 317 LeuValGlyLeuGluGlnGlnGln----- 326
QY 1508 GACCAAGAGAGGAGGGA---AGTAAACAGAGAGGTGCGTAGGTACACAGCGAGGTTGAAG 1564
Db 1508 GACCAAGAGAGGAGGGA---AGTAAACAGAGAGGTGCGTAGGTACACAGCGAGGTTGAAG 1564
QY 327 -----GlyLeuGluSerMetGlnLeuArgArgTyrAlaAlaThrLeuSer 341
Db 327 -----GlyLeuGluSerMetGlnLeuArgArgTyrAlaAlaThrLeuSer 341
QY 1565 GAAGCGCATGTGTTTCATCGCCAGCAGCTCATCCAGTACCATCCAGCTTCCCTCCGAA 1624
Db 1565 GAAGCGCATGTGTTTCATCGCCAGCAGCTCATCCAGTACCATCCAGCTTCCCTCCGAA 1624
QY 342 GluGlyAspIleLeuValIleProSerSerPheProValAlaLeuLysAlaAlaSerAsp 361
Db 342 GluGlyAspIleLeuValIleProSerSerPheProValAlaLeuLysAlaAlaSerAsp 361
QY 1625 CTCATCTGCTTGGCTTCGGTATCAACGCTGAAACCAACACAGAACTTCTCTTCAGGT 1684
Db 1625 CTCATCTGCTTGGCTTCGGTATCAACGCTGAAACCAACACAGAACTTCTCTTCAGGT 1684
QY 362 LeuAsnMetValGlyIleGlyValAsnAlaGluAsnAsnGluArgAsnPheLeuAlaGly 381
Db 362 LeuAsnMetValGlyIleGlyValAsnAlaGluAsnAsnGluArgAsnPheLeuAlaGly 381
QY 1685 GATAAGACAAATGTAGACACGATAGAGAACGAGCGAGGATTTAGCATTCCTCGGG 1744
Db 1685 GATAAGACAAATGTAGACACGATAGAGAACGAGCGAGGATTTAGCATTCCTCGGG 1744
QY 382 AsnLysGluAsnValIleArgGlnIleProArgGlnValSerAspLeuThrPheProGly 401
Db 382 AsnLysGluAsnValIleArgGlnIleProArgGlnValSerAspLeuThrPheProGly 401
QY 1745 TCGGTTGAACAAGTTGAGAGCTCATCAAAACACAGAGGAATCTCCTTTGTGAGTGTCT 1804
Db 1745 TCGGTTGAACAAGTTGAGAGCTCATCAAAACACAGAGGAATCTCCTTTGTGAGTGTCT 1804
QY 402 SerGlyGluGluValGluGluLeuLeuGluAsnGlnLysGluSerTyrPheValAspGly 421
Db 402 SerGlyGluGluValGluGluLeuLeuGluAsnGlnLysGluSerTyrPheValAspGly 421
QY 1805 CGTCTCAA 1813
Db 1805 CGTCTCAA 1813
QY 422 GlnProArg 424
Db 422 GlnProArg 424
```

```
RESULT 14
T44430
protein PVI00 [imported] - winter squash
C:Species: Cucurbita maxima (winter squash)
C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C/Accession: T44430
R/Yamada, K.; Shimada, T.; Kondo, M.; Nishimura, M.; Hara-Nishimura, I.
J. Biol. Chem. 274, 2563-2570, 1999
A:Title: Multiple functional proteins are produced by cleaving Asn-Gln bonds of a single
A:Reference number: 222767; MUID:99107919; PMID:9891029
A/Accession: T44430
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-810 <M>
A/Cross-references: UNIPROT:Q9ZWI3; EMBL:AB019195; NID:G3808061; PIDN:BAA34056.1; PID:G3808061
Alignment Scores:
```

```
Pred. No.: 1,4e-55 Length: 810
Score: 907.50 Matches: 214
Percent Similarity: 54.65% Conservative: 121
Best Local Similarity: 34.91% Mismatches: 191
Query Match: 25.48% Indels: 87
DB: 2 Gaps: 20
US-10-728-051-1 (1-2032) x T44430 (1-810)
```

```
QY 188 CAACAGAAACCGGAT---GACTTGAAGAAAAAGGATGCGAGTCTCGTGCACCAAGCTC 244
Db 188 CAACAGAAACCGGAT---GACTTGAAGAAAAAGGATGCGAGTCTCGTGCACCAAGCTC 244
QY 244 GlnArgAspProAspTyrArgArgGluGlnLysArgArgGluGlnGlnGlnGlnArgArg 263
Db 244 GlnArgAspProAspTyrArgArgGluGlnLysArgArgGluGlnGlnGlnGlnArgArg 263
QY 245 GAGTATGATCTCTGTGTCTATGATCTCGAGGACACACTGGCACCACCAACCAACGCT 304
Db 245 GAGTATGATCTCTGTGTCTATGATCTCGAGGACACACTGGCACCACCAACCAACGCT 304
QY 264 GlnArgGluArgGlyGlyArgAspAspGluAspGlu-----AsnGlnArg 279
Db 264 GlnArgGluArgGlyGlyArgAspAspGluAspGlu-----AsnGlnArg 279
QY 305 TCCCTCCAGGAGGCGGACAGCTGCCGCCCAACCCGAGAC---TACGATGATGACCGC 361
Db 305 TCCCTCCAGGAGGCGGACAGCTGCCGCCCAACCCGAGAC---TACGATGATGACCGC 361
QY 280 AspProAspTyrArgArgGluGlnGluArgArgGluGlnGluArgArgArgGluArg 299
Db 280 AspProAspTyrArgArgGluGlnGluArgArgGluGlnGluArgArgArgGluArg 299
QY 362 CGTCAACCCGAGAGAGAGAGAGGCGCGATGGGACCCAGCTGCGACCCGAGGAGGCTGAA 421
Db 362 CGTCAACCCGAGAGAGAGAGAGGCGCGATGGGACCCAGCTGCGACCCGAGGAGGCTGAA 421
QY 300 GlnGlnGluArgArgGluGluHisArgGlyGlyArgAspValGluAspGluAsnGln 319
Db 300 GlnGlnGluArgArgGluGluHisArgGlyGlyArgAspValGluAspGluAsnGln 319
QY 422 AGAGAAGAAAGACTGGAGACCAACAGAGAGAGATTGGGCGCAGCAAGTATCATCAGCAGCA 481
Db 422 AGAGAAGAAAGACTGGAGACCAACAGAGAGAGATTGGGCGCAGCAAGTATCATCAGCAGCA 481
QY 320 ArgAspProAspTyrArgArgGluGlnGluArgArgGluGluGluGluGluGluGlu 339
Db 320 ArgAspProAspTyrArgArgGluGlnGluArgArgGluGluGluGluGluGluGlu 339
QY 482 CGGAAA---ATAAGGCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 538
Db 482 CGGAAA---ATAAGGCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 538
QY 340 ArgGluTyrGluArgGluHisGlyArgArgGlyArgGluGluGluGluGluGluGlu 359
Db 340 ArgGluTyrGluArgGluHisGlyArgArgGlyArgGluGluGluGluGluGluGlu 359
QY 539 CATGTGAGG----- 547
Db 539 CATGTGAGG----- 547
QY 360 GlnArgArgHisGluArgGlnHisGlyArgSerArgValAsnGlnValAlaIle 379
Db 360 GlnArgArgHisGluArgGlnHisGlyArgSerArgValAsnGlnValAlaIle 379
QY 548 -----GAAGAAACATCTCGGAACCAACCTTCTTACTTCCCGTCAAGGCGGTTAGC 598
Db 548 -----GAAGAAACATCTCGGAACCAACCTTCTTACTTCCCGTCAAGGCGGTTAGC 598
QY 380 ArgArgThrGluGlnGluGlnSerAsnAsnProTyrTyrPheGlnGlnGlnArgPheGln 399
Db 380 ArgArgThrGluGlnGluGlnSerAsnAsnProTyrTyrPheGlnGlnGlnArgPheGln 399
QY 599 ACCCGCTAGGGAACCAAAACGGTAGGATCCGGTCTCGAGAGGTTTGACCAAGGTCA 658
Db 599 ACCCGCTAGGGAACCAAAACGGTAGGATCCGGTCTCGAGAGGTTTGACCAAGGTCA 658
QY 400 SerArgTyrArgSerAspGluGlyHisTyrArgValLeuGluArgPheSerGluArgSer 419
Db 400 SerArgTyrArgSerAspGluGlyHisTyrArgValLeuGluArgPheSerGluArgSer 419
QY 659 AGGCAGTTTCAGAACTCCAGAACTCCAGTATTCGAGATCGAGGCGCAAAACCTTAACACT 718
Db 659 AGGCAGTTTCAGAACTCCAGAACTCCAGTATTCGAGATCGAGGCGCAAAACCTTAACACT 718
QY 420 GlnLeuLeuLysGlyLysAsnGlnArgLeuAlaLeuLeuGluAlaArgProHisThr 439
Db 420 GlnLeuLeuLysGlyLysAsnGlnArgLeuAlaLeuLeuGluAlaArgProHisThr 439
QY 719 CTTGTTCTTCCCAAGCAGCTGATGCTGATTAACATCTTGTATTCAGCAAGGCGCAAGCC 778
Db 719 CTTGTTCTTCCCAAGCAGCTGATGCTGATTAACATCTTGTATTCAGCAAGGCGCAAGCC 778
QY 440 PheIleValProHisHisLeuAspAlaGluCysValLeuLeuValValArgGlyArgAla 459
Db 440 PheIleValProHisHisLeuAspAlaGluCysValLeuLeuValValArgGlyArgAla 459
QY 779 ACCGTGACCGTAGCA-----AATGCGAATAACAGAAAGAGAGCTTTAATCTTGAC 826
Db 779 ACCGTGACCGTAGCA-----AATGCGAATAACAGAAAGAGAGCTTTAATCTTGAC 826
QY 460 ThrIleThrThrValValGlnGluLysArgGluThrArgLysGluSerTyrAsnValGlu 479
Db 460 ThrIleThrThrValValGlnGluLysArgGluThrArgLysGluSerTyrAsnValGlu 479
QY 827 GAGGCGCATGCACTCAGAAATCCCATCCGTTTTCATTTCTTCTACATCTTGAACCGCATGAC 886
Db 827 GAGGCGCATGCACTCAGAAATCCCATCCGTTTTCATTTCTTCTACATCTTGAACCGCATGAC 886
QY 480 SerGlyAspValMetThrIleProAlaGlyThrThrLeuTyrLeuAlaAsnGln---Glu 498
Db 480 SerGlyAspValMetThrIleProAlaGlyThrThrLeuTyrLeuAlaAsnGln---Glu 498
QY 887 AACCAAGACCTCAGAGTAGCTTAAATCTCCATGCCGTTTAAACACACCGCGCAGTTTGAG 946
Db 887 AACCAAGACCTCAGAGTAGCTTAAATCTCCATGCCGTTTAAACACACCGCGCAGTTTGAG 946
QY 499 AsnGluAspLeuGlnIleValLysLeuValGlnProValAsnAsnProGlyGluPheLys 518
Db 499 AsnGluAspLeuGlnIleValLysLeuValGlnProValAsnAsnProGlyGluPheLys 518
QY 947 GATTCTTCCCGCGGAGCAGCGGAGAGCAATCATCTTCTTCTGAGGCGCTTCCAGGAGAT 1006
Db 947 GATTCTTCCCGCGGAGCAGCGGAGAGCAATCATCTTCTTCTGAGGCGCTTCCAGGAGAT 1006
QY 519 AspTyrLeuSerAlaGlyGlyGluSerGln---AlaTyrTyrSerValPheSerAsnAsp 537
Db 519 AspTyrLeuSerAlaGlyGlyGluSerGln---AlaTyrTyrSerValPheSerAsnAsp 537
QY 1007 ACGTTGGAGCGCGCTTCAATATCGGAAATTCATGAGATACGAGGCGTCTGTTAGAGAGAG 1066
Db 1007 ACGTTGGAGCGCGCTTCAATATCGGAAATTCATGAGATACGAGGCGTCTGTTAGAGAGAG 1066
QY 538 ValLeuGluAlaAlaLeuAsnIleProArgAspLysLeuGluArgIlePhe----- 554
Db 538 ValLeuGluAlaAlaLeuAsnIleProArgAspLysLeuGluArgIlePhe----- 554
```


1067	Qy	AATCCAGGAGGTGACGAAGAGGAGAGAGGCGAGGCGATGGAGTACTCGGAGTAGTGTGAG	1120
555	Db	-----LysGlnArgfGluArgGly-----	561
1127	Qy	AACAATGAAGGAGTGATGACCAAGTCTCAAGAGGACGACGTTGAAGAACTTACTAAGCAC	1186
562	Db	-----GlyLysIleValArgGalaserGlnGluLeuArgAlaLeuSerGlnArg	578
1187	Qy	GCTAAATCCGTCTCAAGAAAGGCTCCGAAGAGGAGGATATACCAACCCCAATCAAC	1246
579	Db	AlaThrSerVal---ArgLysGlySerArg-----GlyValArgAlaProIleLys	594
1247	Qy	TTGAGAGAAGCGAGCCGATCTTCTTCAACACTTTGGAGAGTTATTGAGGTGAAGCCA	1306
595	Db	LeuGluSerGlnThrProValTyAsnAsnGlnTyGlyGlnMetPheGluAlaCysPro	614
1307	Qy	GACAAGAAGAACCCAGCTTCAGGACCTGGACATGATGTCACCTCTGTAGAGATCAAA	1366
615	Db	Asp---GluPheProGlnLeuArgArgThrAspValAlaThrSerValValaspIleLys	633
1367	Qy	GAAGGAGCTTGTGCTCCACACTTCAACTCAAAAGGCCATGGTTATCTCGTCGTCAAC	1426
634	Db	GlnGlyMetMetValProHisPheAsnSerArgAlaThrTrpValValPheValSer	653
1427	Qy	AAAGGAATCGAAACCTTGAACCTCGTGGCT-----GTAAAGAAAGAGCAACACAGAGG	1480
654	Db	GluGlyAlaGlySerPheGluMetAlaCysProHisIleGlnSerSerGlnTrpGlnArg	673
1481	Qy	GGACGGCGGNAGNAGAG-----GAGACGACGACGAAAGAGGAGGGA	1525
674	Db	GlyArgArgGluGluGluArgHisTrpArggGluGluGluGluArgGluGluArg	693
1526	Qy	AGTAACACAGAGGTGCGTAGGTACACAGCAGGTTTGAAGAGCGCATGTTTCATCATG	1585
694	Db	SerGlyArg---PheGluArgValAlaGlyArgLeuSerGluGlyGlyValLeuValIle	712
1586	Qy	CCAGCAGCTCATCCAGTAGCCATCAACGCTTCC-----TCCGAACCTCCATCTGCTTGGC	1639
713	Db	ProAlaGlyHisProIleAlaIleMetAlaSerProAsnGluAsnLeuArgLeuValGly	732
1640	Qy	TTGCGTATCAACCGCTGAAACCAACACAGAAATCTTCTTCGAGGTGATAGGACAATGTG	1699
733	Db	PheGlyIleAsnAlaGluAsnAsnHisArgAsnPheLeuAlaGly---ArgGluAsnIle	751
1700	Qy	ATAGACCAGATAGAGAAGCAAGCAAGGNTTAGCATTCCTCGGTCCGTCGTCGAACAGTT	1759
752	Db	MetAsnGluLeuAspArgGluAlaLysGluLeuAlaPheAsnValGluGlyLysGlnAla	771
1760	Qy	GAGAAGCTCATCAAAAACCAAGAGGAATCTCACTTTGTGAGTGTCTGCTCCTCAATCTCAA	1819
772	Db	AspGluIlePheArgSerGlnArgGluSerPhePhe-----	783
1820	Qy	TCTCAATCTCCGTGCTTCTCTGAGAAGAGTCTCCTCGAAGAGGATCAAGAGGAGGAA	1879
784	Db	ThrGluGlyProGluGlyGlyArgArgArgSerThrGlu-----	796
1880	Qy	AACCAAGGAGGAGGTCCACTCTCTTCAATTTTGAAG	1918
797	Db	-----ArgSerProLeuLeuSerIleLeuLys	805

RESULT 15

S06398
alpha-globulin type A precursor - upland cotton
N:Alternate names: seed storage protein
C:Species: *Gossypium hirsutum* (upland cotton)
C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 09-Jul-2004
C:Accession: S06398
R:Chlan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
Plant Mol. Biol. 9, 533-546, 1987
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX. Sequence determination of the alpha-globulin type A precursor
A:Reference number: S06398
A:Accession: S06398
A:Status: not compared with conceptual translation

A;Molecule type: DNA
A;Residues: 1-605 <CHL>
A;Cross-references: UNIPROT:P09799
C;Superfamily: glycinin
F;1-34/domain: signal sequence #status predicted <SIG>
F;25-605/product: alpha-globulin type A #status predicted <MAT>

Alignment Scores:		
Pred. No.:	8.2e-51	Length:
Score:	839.00	Matches:
Percent Similarity:	50.86%	Conservative:
Best Local Similarity:	33.91%	Mismatches:
Query Match:	23.55%	Indels:
DB:	2	Gaps:
		21
		605

US-10-728-051-1 (1-2032) x S06398 (1-605)

Qy	65	TCTCCACTGATGCTGTGTTGCTTAGGGAGTCTTGTCTCGGCTTCAAGTTCTTCTGCAACGATGCC	124
Db	6	SerValPheValValLeuPheSerLeuPheLeuSerPheGlyLeuLeuCysSerAla	25
Qy	125	AAGTCATCATCCTTACCAGAAACACAGAACCCCTCGCCCCAGAGG	175
Db	26	LysaspPheProGlyArgArgSerGluaspAspPro---GlnGlnArgTyrGluaspCys	44
Qy	176	CTCCAGAGTTGTC AACAGGAACCG---GATGACTTTGAAGCAAAAGGCATGCGAGTCTCGC	232
Db	45	ArgLysArgCysGlnLeuGluThrArgGlyGlnThrGluGlnaspLysCysGluaspArg	64
Qy	233	TGC---ACCAAGCTCGAG-----TATGATCTCGTTGTGTCTTAT	268
Db	65	SerGluThrGlnLeuLysGluGlnGlnArgAspGlyGluaspProGlnArgArgTyr	84
Qy	269	---GATCTCTGAGGACACATCGGCACCACCAACCACTGCCCTCCAGGG-----	316
Db	85	GlnaspCysArgGlnHisCysGlnGlnGluArgArgLeuArgProHisCysGluGln	104
Qy	317	-----GAGCGGACACGTGGCCGCCAACCCGGAGACTACGATGATGAC	358
Db	105	SerCysArgGluGlnTyrGluLysGlnGlnGlnGlnPro-----Asp	119
Qy	359	CGCGCTCAACCCGAGAGAGGAGGAGCGCGATGGGGACCAGCTGGACCG-----	409
Db	120	LysGlnPheLysGluCysGlnGlnArgCysGlnTrpGlnGlnArgProGluArgLys	139
Qy	410	-----AGGAGCGCTGAAAGAGAAAGACTGTGAGACAAACCAAGA	448
Db	140	GlnGlnCysValLysGluCysArgGluGlnTyrGlnGluaspProTrpLysGlyGluArg	159
Qy	449	GAAGAT---TGGAGCGGACCAAGTCATCAGCAGCCACCGGAATTAAGGCCCCGAAAGGA	505
Db	160	GluAsnLysTrpArgGluGluGluGlu-----GluSerAsp	172
Qy	506	GAAGGAGAAACAAGAGTGGGGAACACACAGGTAGCATGTGAGGGAAGAAACATCTCGGAAC	565
Db	173	GluGlyGluGlnGln-----GlnArgAsn	180
Qy	566	AACCTTTTCTACTTCCCGTCAAGCGGTTTAGCACCCCGCTACGGGAACCAAAACCGTAGG	625
Db	181	AsnProTyrTyrPheHisArgSerPheGlnGluArgPheArgGluGluHisGlyAsn	200
Qy	626	ATCCGGGTCTGCAGAGGTTTGACCAAAGGTCAAGGCAGTTTCAGATCTCCAGATCAC	685
Db	201	PheArgValLeuGlnArgPheAlaAspLysHisLeuLeuArgGlyIleAsnGluPhe	220
Qy	686	CGTATTGTGCATGTCGAGGCCAAACCTAACACTCTTGTCTTCCCAAGCACGCTGATGCT	745
Db	221	ArgIleAlaIleLeuGluAlaAsnProAsnThrPheValLeuProHisCysaspAla	240
Qy	746	GATAACATCTCTGTTATCCAGCAGGCGCAACCGTCAGCCGTAGCAANTGGCAATAAC	805
Db	241	GluLysIleTyrValValThrAsnGlyArgGlyThrValThrPheValThrHisGluAsn	260

A/Status: not compared with conceptual translation

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 8, 2005, 03:45:11 ; Search time 343.105 Seconds
(without alignments)
6065.459 Million cell updates/sec

Title: US-10-728-051-1
Perfect score: 3562
Sequence: 1 aataacatatatttcac.....cgtttgctgtttcttc 2032

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 3224756

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2 1/USPTO.spool_p/US10728051/runat_07042005_125045_14049/app.query.fasta_1.4757
-DB=UniProt -QFMT=fastaan -SUFFIX=rup -MINMATCH=0.1 -LOOPCT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEA=SIZE=500 -MINLEN=0 -MAXLEN=200000000
-USR=US10728051@CGN 1.1 874 @runat_07042005_125045_14049 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt_03.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3286	92.3	626	1 AL12 ARAHY	P43238 arachis hyp
2	3041	85.4	614	1 AL11 ARAHY	P43237 arachis hyp
3	2887.5	81.1	580	2 Q6PSU3	Q6psu3 arachis hyp
4	2172.5	61.0	428	2 Q6PSU4	Q6psu4 arachis hyp
5	1495.5	42.0	299	2 Q6PSU5	Q6psu5 arachis hyp
6	1435	40.3	303	2 Q6PSU6	Q6psu6 arachis hyp
7	1302	36.6	621	2 Q7XHT2	Q7xht2 glycine max
8	1290	36.2	533	2 Q6EBG1	Q6ebcl lupinus alb
9	1284	36.0	621	2 Q948X0	Q948y0 glycine max
10	1282	36.0	605	2 Q94LX2	Q94lx2 glycine max
11	1281	36.0	605	1 GLCA SOYBN	P13916 glycine max
12	1267	35.6	571	1 CVCA_PEA	P13915 pisum sativ
13	1265.5	35.5	623	2 Q948X9	Q948x9 glycine max
14	1258.5	35.3	545	2 Q41674	Q41674 vicia narbo
15	1251	35.1	613	2 Q9M3X6	Q9m3x6 pisum sativ
16	1251	35.1	639	1 GLCX SOYBN	P11827 glycine max

17	1190	33.4	559	2 Q9F2P9	Q9fzp9 glycine max
18	1183.5	33.2	418	2 Q84U11	Q84u11 lens culina
19	1180.5	33.1	543	2 Q22120	Q22120 glycine max
20	1178.5	33.1	459	1 VCLC_PEA	P13918 pisum sativ
21	1168	32.8	415	2 Q84U10	Q84u10 lens culina
22	1165	32.7	416	2 Q22121	Q22121 glycine max
23	1156	32.5	439	1 GLCB_SOYBN	P25974 glycine max
24	1152	32.3	439	2 Q93VL9	Q93vl9 glycine max
25	1148	32.2	438	2 Q43626	Q43626 pisum sativ
26	1145	32.1	415	2 Q702P1	Q702p1 pisum sativ
27	1135	31.9	415	2 Q702P0	Q702p0 pisum sativ
28	1123.5	31.5	453	2 Q41677	Q41677 vicia narbo
29	1108.5	31.1	483	1 VCL_VICFA	P08438 vicia faba
30	1082	30.4	396	2 Q84UB3	Q84ub3 glycine max
31	1068	30.0	410	1 VCLB_PEA	P02854 pisum sativ
32	1030	28.9	518	2 Q9M3X8	Q9m3x8 lens culina
33	962.5	27.0	445	1 CANA CANEN	P50477 canavalia e
34	959.5	26.9	445	1 CANA_CANGL	P10562 canavalia g
35	910	25.5	593	2 Q9SEW4	Q9sew4 juglans reg
36	907.5	25.5	810	2 Q9ZWI3	Q9zwi3 cucurbita m
37	880	24.7	481	2 Q7Y1C1	Q7y1c1 juglans nig
38	865.5	24.3	666	2 Q9SPL4	Q9spl4 macadamia i
39	863.5	24.2	625	2 Q9SPL3	Q9spl3 macadamia i
40	849.5	23.8	666	2 Q9SPL5	Q9spl5 macadamia i
41	839	23.6	605	1 VCLA_GOSHI	P09799 gossypium h
42	825.5	23.2	436	1 PHSA_PHAVU	P07219 phaseolus v
43	806.5	22.6	430	2 Q41115	Q41115 phaseolus v
44	805.5	22.6	430	2 Q43633	Q43633 phaseolus v
45	804	22.6	423	1 PHS2_PHALU	Q43617 phaseolus l

ALIGNMENTS

RESULT 1

AL12 ARAHY
ID AL12 ARAHY STANDARD; PRT; 626 AA.
AC P43238;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Allergen Ara h 1, clone P41B precursor (Ara h I).
OS Arachis hypogaea (Peanut).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;
OC Arachis.
OX NCBI_TaxID=3818;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Florunner;
RX MEDLINE=96013631; PubMed=7560062;
RA Burks A.W., Cockrell G., Stanley J.S., Helm R.M., Bannon G.A.;
RT "Recombinant peanut allergen Ara h I expression and IGE binding in
RT patients with peanut hypersensitivity.";
RL J. Clin. Invest. 96:1715-1721(1995).
RN [2]
RP CARBOHYDRATE-LINKAGE SITE ASN-521.
RX MEDLINE=20455243; PubMed=10998264; DOI=10.1006/abio.2000.4737;
RA Kolarich D., Altmann F.;
RT "N-glycan analysis by matrix-assisted laser desorption/ionization mass
RT spectrometry of electrophoretically separated nonmammalian proteins:
RT application to peanut allergen Ara h I and olive pollen allergen Ole e
RT 1.";
RL Anal. Biochem. 285:64-75(2000).
CC -I- ALLERGEN. Causes an allergic reaction in human.
CC -I- SIMILARITY: Belongs to the 7S seed storage protein family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/

```

or send an email to license@isb-sib.ch).
-----
CC EMBL; L34402; AAB00861.1; -.
DR HSSP; P25974; 1IPJ.
DR InterPro; IPR006045; Cupin.
DR InterPro; IPR007113; Cupin region.
DR InterPro; IPR011051; RmlC_like_cupin.
DR Pfam; PF00190; Cupin; 2.
KW Allergen; Glycoprotein; Signal.
FT SIGNAL 1 25 potential.
FT CHAIN 26 626 Allergen Ara h 1, clone P41B.
FT CARBOHYD 521 521 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 626 AA; 71345 MW; 1A6BBE41490D0E3 CRC64;

Alignment Scores:
Pred. No.: 2,34e-207 Length: 626
Score: 3286.00 Matches: 626
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 92.25% Indels: 0
DB: 1 Gaps: 0

US-10-728-051-1 (1-2032) x AL12_ABAHY (1-626)
QY 50 ATGAGAGGAGGGTTTCTCAGTATGCTGCTTCTAGGATCCTTGTCTCGGCTTCAGTT 109
DB 1 MetArgGlyArgValSerProLeuMetLeuLeuGlyLeuValLeuAlaSerVal 20
QY 110 TCTGCAACGATCCCAAGTCATCACCTTACCAGAAGAAACAGAGAACCCCTGCGCCAG 169
DB 21 SerAlaThrHisAlaLysSerProTyrGlnLysThrGluAsnProCysAlaGln 40
QY 170 AGTGCTCTCAGAGTTGTCACAGGAACCGGATGACTTGAAGCAAAAGGCATCGAGTCT 229
DB 41 ArgCysLeuGlnSerCysGlnGlnProAspAspLeuLysGlnLysAlaCysGluSer 60
QY 230 CGTGCACCAAGCTCAGATATGATCTCTGTTGTCTATGATCTCTCGAGGACACTGGC 289
DB 61 ArgCysThrLysLeuGluTyrAspProArgCysValTyrAspProArgGlyHisThrGly 80
QY 290 ACCACCAACCAAGTTCCTCCCTCAGGGGAGCGACACGTCGGCGCCAAACCGGAGACTAC 349
DB 81 ThrThrAsnGlnArgSerProGlyGluArgThrArgGlyArgGlnProGlyAspTyr 100
QY 350 GATGATCACCGCGTCAACCCCGAAGAGAGAGAGGCGGATGGGACCGAGCTGGACCG 409
DB 101 AspAspArgArgGlnProArgArgGluGluGlyArgTyrGlyProAlaGlyPro 120
QY 410 AGGAGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 469
DB 121 ArgGluArgGluArgGluArgGluArgGluArgGluArgGluArgGluArgGluArg 140
QY 470 CATCAGCAGCCAGGAATAAGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 529
DB 141 HisGlnGlnProArgLysIleArgProGluGlyArgGluGlyGluGlnGluTyrGly 160
QY 530 CCAGGTAGCCATGTGAGGGAAGAAACATCTCGGAACAACCCCTTCTACTTCCCGTCAAG 589
DB 161 ProGlySerHisValArgGluGluThrSerArgAsnAsnProPheTyrPheProSerArg 180
QY 590 CGGTTTAGCACCCTGACGGGAACCAACCGTAGGATCCGGTCTCTCGAGGTTTGAC 649
DB 181 ArgPheSerThrArgTyrGlyAsnGlnAsnGlyArgIleArgValLeuGlnArgPheAsp 200
QY 650 CAAAGGTCAAGGAGGTTTCAAGATCTCCAGAACTCCAGAACTCCGATTCGAGATCGAGGCCAA 709
DB 201 GlnArgSerArgGlnPheGlnAsnLeuGlnAsnHisArgIleValGlnIleGluAlaLys 220
QY 710 CCTAACACTCTTCTCCAGGACCGTGTGCTGATCAATCTCTGATCAATCTCTGAGAAAGAG 769
DB 221 ProAsnThrLeuValLeuProLysHisAlaAspAlaAspAsnIleLeuValIleGln 240
QY 770 GGCGAAGCCACCGTACCGCTAGCAATGGCAATAACAGAAAGAGCTTTAATCTTGACGAG 829

```

```

241 GlyGlnAlaThrValThrValAlaAsnGlyAsnAsnArgLysSerPheAsnLeuAspGlu 260
830 GGCCATGCACCTCAGAAATCCCATCCCGTTTTCATTTCTACATCTTGAACCGCCATGACAA 889
261 GlyHisAlaLeuArgIleProSerGlyPheIleSerTyrIleLeuAsnArgHisAspAsn 280
890 CAGAACTCAGAGTAGCTAAATCTCCATGCCGTTAAACACACCCGCGCAGTTTGAAGAT 949
281 GlnAsnLeuArgValAlaLysIleSerMetProValAsnThrProGlyGlnPheGluAsp 300
950 TTCTTCCCGCCAGCAGCCGAGACCAATCATCTTCTTCCAGGGCTTCAGCAGGAATAG 1009
301 PhePheProAlaSerSerArgAspGlnSerSerTyrLeuGlnGlyPheSerArgAsnThr 320
1010 TTGGAGGCGCTTCAATCGCGAAATTCATAGATACGAGGGTGTGTTTAGAAGAGAA 1069
321 LeuGluAlaAlaPheAsnAlaGluPheAsnGluIleArgArgValLeuLeuGluAsn 340
1070 GCAGGAGGTGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1129
341 AlaGlyGlyGluGlnGluArgGlyGlnArgArgTyrSerThrArgSerSerGluAsn 360
1130 AATGAGAGAGTATAGTCAAGTGTCAAGGAGGACGTTGAAGAACTTACTTAAGCAGCT 1189
361 AsnGluGlyValIleValLysValSerLysGluHisValGluGluLeuThrLysHisAla 380
1190 AAATCGCTCTCAAGAAAGCGTCCGAAGAGAGAGAGAGATATCACCAACCAATCAACTTG 1249
381 LysSerValSerLysLysGlySerGluGluGluGlyAspIleThrAsnProIleAsnLeu 400
1250 AGAGAAGCGAGCCCGATCTTTTAACTTTGGAAAGTATTTAGGTGAAGCCAGAC 1309
401 ArgGluGlyGluProAspLeuSerAsnAsnPheGlyLysLeuPheGluValLysProAsp 420
1310 AGAGAAGACCCCGAGCTTCAGCACTGAGCATGATGCTTCACTGTGTGAGATCAAGAA 1369
421 LysLysAsnProGlnLeuGlnAspLeuAspMetLeuThrCysValGluLysGlu 440
1370 GGAGCTTTTATGCTCTCCACACTTCAACTCAAGGCGCATGTTTATCGTCTCGTCAACAA 1429
441 GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValAlaAsnLys 460
1430 GGAATCGGAAACCTTGAATCTGTGCTGTAAGAAAGAGCAACAACAGAGGGAGCGCGG 1489
461 GlyThrGlyAsnLeuGluValAlaValArgLysGluGlnGlnArgGlyArg 480
1490 GAAGAGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTAC 1549
481 GluGluGluGluAspGluAspGluGluGluGluGlySerAsnArgGluValArgArgTyr 500
1550 ACAGCGAGGTTCAAGGAAGCGATGTTTCATCATGCGCAGCAGCTCATCCAGTAGCCATC 1609
501 ThrAlaArgLeuLysGluGlyAspValPheIleMetProAlaAlaHisProValAlaIle 520
1610 AACGTTCTCCGAACTCCATCTGTGCTTGGCTTCGGTATCAACGCTGAAACCAACACAGA 1669
521 AsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsnHisArg 540
1670 ATCTTCTTTCAGGTGATAGGACATGATGTAGACCATAGAGAGAGAGAGAGAGAGAT 1729
541 IlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGlnAlaLysAsp 560
1730 TTAGCATTCCTCGGTCGGGTGAACAAAGTTGAGAGCTCATCAAAACAGAGAGAGATCT 1789
561 LeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGlnLysGluSer 580
1790 CACTTTGTGAGTGTCTCTCTCAATCTCAATCTCAATCTCGTCTCTCTGAGAAAGAG 1849
581 HisPheValSerAlaArgProGlnSerGlnSerProSerSerProGluLysGlu 600
1850 TCTCTCGAAGAGAGGATCAAGAGGAGGAGAAACCAAGGAGGAGGAGGAGGAGGAGGAG 1909

```

Db 601 SerProGluLysGluAspGlnGluGluAsnGlnGlyGlyLysGlyProLeuLeuSer 620

Qy 1910 ATTTTGAAGCTTTTAAAC 1927

Db 621 IleLeuLysAlaPheAen 626

RESULT 2

AL11_ARAHY STANDARD; PRT; 614 AA.

AC P43237;

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 11-NOV-1995 (Rel. 32, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Allergen Ara h 1, clone P17 precursor (Ara h I).

OS Arachis hypogaea (Peanut).

OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;

OC Arachis.

OX NCBI_TaxID=3818;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Florunner;

RX MEDLINE=96013631; PubMed=7560062;

RA Burks A.W., Cockrell G., Stanley J.S., Helm R.M., Bannon G.A.;

RT "Recombinant peanut allergen Ara h I expression and IgE binding in

RT patients with peanut hypersensitivity.";

RL J. Clin. Invest. 96:1715-1721(1995).

RN [2]

RP CARBOHYDRATE-LINKAGE SITE ASN-516.

RX MEDLINE=20455243; PubMed=1098264; DOI=10.1006/abio.2000.4737;

RA Kolarich D., Altmann F.;

RT "N-glycan analysis by matrix-assisted laser desorption/ionization mass

RT spectrometry of electrophoretically separated nonmammalian proteins:

RT application to peanut allergen Ara h 1 and olive pollen allergen Ole e

RT 1.";

RL Anal. Biochem. 285:64-75(2000).

CC -/- ALLERGEN: Causes an allergic reaction in human.

CC -/- SIMILARITY: Belongs to the 7S seed storage protein family.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC -----

DR ENBL; L38853; AAA60336.1; -.

DR HSSP; P25974; IIPJ.

DR GlycoSuiteDB; P43237; -.

DR InterPro; IPR006045; Cupin.

DR InterPro; IPR007113; Cupin region.

DR InterPro; IPR011051; RmC_Like_cupin.

DR Pfam; PF00190; Cupin; 2.

KW Allergen; Glycoprotein; Signal.

FT SIGNAL 1 25 Potential.

FT CHAIN 26 614 Allergen Ara h 1, clone P17.

FT CARBOHYD 516 516 N-linked (GlcNAc...).

FT SEQUENCE 614 AA; 70283 MW; 1DDCF217EBC5F31 CRC64;

Alignment Scores:

Pred. No.:	2.9e-191	Length:	614
Score:	3041.00	Matches:	595
Percent Similarity:	96.03%	Conservative:	9
Best Local Similarity:	94.59%	Mismatches:	7
Query Match:	85.37%	Indels:	18
DB:	1	Gaps:	7

US-10-728-051-1 (1-2032) x AL11_ARAHY (1-614)

Qy 50 ATGAGAGGAGGGTTTCTCCATGATGCTGTGCTAGGGATCCTTGCTCGCTTCAGTT 109

Db	1	MetArgGlyArgValSerProLeuMetLeuLeuGlyIleLeuValLeuAlaSerVal	20
Qy	110	TTGTCAACGCATGCCAAGTCATCACCTTACAGAGAAAAACAGAGAACCCCTCGCCACG	169
Db	21	SerAlaThrGlnAlaLys---SerProTyr---ArgLysThrGluAenProCysAlaGln	38
Qy	170	AGGTGCTCCAGAGTTGTCTCAACAGAACCGGATGACTTGAAGCAAAAGGATCGAGTCT	229
Db	39	ArgCysLeuGlnSerCysGlnGluProAspAspLeuLysGlnLysAlaCysGluSer	58
Qy	230	CGCTGCACCAAGCTCCAGTATGATCTCTGTTGTCTATGATCTCGAGGACACACATGCC	289
Db	59	ArgCysThrLysLeuGluTyrAspProArgCysValTyrAsp-----ThrGly	74
Qy	290	ACCACCAACCAACGTTCCCTCCAGGGGAGCGACACGTCGCGCCCAACCCGGAGACTAC	349
Db	75	AlaThrAsnGlnArgHisProProGlyGluArgThrArgGlyArgGlnProGlyAspTyr	94
Qy	350	GATGATGACCGCTCAACCCCGAAGAGAGAGAGGCGCGATGGGGACCCAGCTGGACCG	409
Db	95	AspAspAspArgArgGlnProArgArgGluGluGlyGlyArgTyrProAlaGluPro	114
Qy	410	AGGAGCGGTGAAGAGAGAGAGACTCGAGACACCAAGAGAGAGATTCGAGCGCACCACT	469
Db	115	ArgGluArgGluArgGluGluAspTyrArgGlnProArgGluAspTyrArgProSer	134
Qy	470	CATCAGCAGCCCGGAAATAAGCCCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	529
Db	135	HisGlnGlnProArgLysIleArgProGluGlyArgGluGluGluGluTyrGlyThr	154
Qy	530	CAAGTAGCATGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	589
Db	155	ProGlySerGluValArgGluThrSerArgAsnAsnProPheTyrPheProSerArg	174
Qy	590	CGGTTTAGCACCCTGACCGGACCAAAACGGTAGGATCCGGCTCGGAGAGGTTTAC	649
Db	175	ArgPheSerThrArgTyrGlyAsnGlnAsnGlyArgIleArgValLeuGlnArgPheAsp	194
Qy	650	CAAAGTCAAGGAGAGTTTCAAGATCCAGAAATCCAGCTATTCTGCAGATCGAGGCCAA	709
Db	195	GlnArgSerLysGlnPheGlnAsnLeuGlnAsnHisArgIleValGlnIleGluAlaArg	214
Qy	710	CTTAAACATCTTGTCTTCCAAAGCAGCTGATGATGATGATGATGATGATGATGATG	769
Db	215	ProAsnThrLeuValLeuProLysHisAlaAspAlaAspAsnIleLeuValIleGlnGln	234
Qy	770	GGCAGACCCAGCTGACCGCTAGCAATGCAATGCAATGCAATGCAATGCAATGCAATG	829
Db	235	GlyGlnAlaThrValThrValAlaAlaAsnGlyAsnAsnArgLysSerPheAsnLeuAspGlu	254
Qy	830	GGCCATGCTCATCAGAAATCCCATCCGTTTTCATTTCTTCTTCTTCTTCTTCTTCTT	889
Db	255	GlyHisAlaLeuArgIleProSerGlyPheIleSerTyrIleLeuAsnArgHisAspAsn	274
Qy	890	CAGAACCTCAGAGTAGCTAAATCTCCATGCGCGTTTAAACACACCCGGCCAGTTTGGAT	949
Db	275	GlnAsnLeuArgValAlaLysIleSerMetProValAsnThrProGlyGlnPheGluAsp	294
Qy	950	TTCTTCCCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG	1009
Db	295	PhePheProAlaSerSerArgAspGlnSerSerTyrLeuGlnGlyPheSerArgAsnThr	314
Qy	1010	TTGGAGCGCGCTTCAATCGGAATTCATGAGATACGAGGAGGAGGAGGAGGAGGAGG	1069
Db	315	LeuGluAlaAlaPheAsnAlaGluPheAsnGluIleArgArgValLeuLeuGluAlaAsn	334
Qy	1070	GCAGGAGGTGAGCAAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1129
Db	335	AlaGlyGlyGluGlnGluGluArgGlyGlnArgArgSerThrArgSerSerAsp---	353
Qy	1130	AATGAGGAGTGTAGTCAAAGTGTCAAAGGAGCGGTTGAAGACTTACTTAAGCAGCT	1189

Qy	890	CAGAACCTCAGAGTAGCTAAATCTCCATGCCCGTTHAACACACCCGCCAGTTTGGAGAT	949
Db	275	GlnAsnLeuArgValAlaLysIleSerMetProValAsnThrProGlyGlnPheGluAsp	294
Qy	950	TTCTTCCCGCGCAGCAGCCAGACCAATCATCTTACCTTGCAGGGCTTCACAGGMAATACG	1009
Db	295	PhePheProIleSerSerArgAspGlnSerSerTyrLeuGlnGlyPheSerArgAsnThr	314
Qy	1010	TTGAGGCCCGCTTCAATGCGGAATTCATGAGATACGAGGGTGCCTGTAGAGAGAAT	1069
Db	315	LeuGluAlaAlaPheAsnAlaGluPheAsnGluIleArgArgValLeuLeuGluAsn	334
Qy	1070	GCAGGAGGTGAGCAGAGAGAGAGGCGCAGAGCGGATGAGATCTCGAGTAGTGAGAAC	1129
Db	335	AlaGlyGlyGluGlnGluGluArgGlyGlnArgArgSerThrArgSerSerAsp---	353
Qy	1130	AATCGAGGAGCTGATAGTCAAGTGTCTAAAGGACGACGCTTCAAGAACTTACTAGACCGCT	1189
Db	354	AsnGluGlyValIleValIleValIleValSerLysGluHisValGlnGluLeuThrLysHisAla	373
Qy	1190	AAATCCGCTCAAAAGAAAGGCTCCGAAGAAGAGGGAGATATCACCAACCCAACTCAACTG	1249
Db	374	LysSerValSerLysLysGlySerGluGluGlu---AspIleThrAsnProIleAsnLeu	392
Qy	1250	AGAAAGGCGAGCCCGATCTTCTTAAACAACCTTTGGGAAGTTATTGTAGGTGAAGCCAGAC	1309
Db	393	ArgAspGlyGluProAspLeuSerAsnAsnPheGlyArgLeuPheGluValLysProAsp	412
Qy	1310	AAGAAGAACCCCGAGCTTCAGGACCTGGACATGATGCTCACTGTGTAGAGATCAAGAA	1369
Db	413	LysLysAsnProGlnLeuGlnAspLeuAspMetMetLeuThrCysValGluIleLysGlu	432
Qy	1370	GGAGCTTTGATGCTCCACACTTCAACTCAAAAGGCCATGTTATCGTCGTCGCAACAA	1429
Db	433	GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValAsnLys	452
Qy	1430	GGAACTGGAAACCTTGAACCTCGTGGCTGTGAAGAAAGAGCAACACAGAGGGGACCGCGG	1489
Db	453	GlyThrGlyAsnLeuGluLeuValAlaValAlaGlyGluGlnGlnGlnArgGlyArgArg	472
Qy	1490	GAA-----GAAGAGGAGGACGAAGACGAAGAGAGGAGGAGGAGTAACAGAGAGGTG	1540
Db	473	GluGlnGluTrpGluGluGluGluGluAspGluGluGluGluGlySerAsnArgGluVal	492
Qy	1541	CGTAGGTACACAGCGAGGTTGAAGGAGGCGATGTTTCATCATGCCAGCAGCTCATCCA	1600
Db	493	ArgArgTyrThrAlaArgLeuLysGluGlyAspValPheIleMetProAlaAlaHisPro	512
Qy	1601	GTAGCCATCAACGCTTCCTCCGAACCTCCATCTGCTTGGCTTCGGTATCAACGCTGAAAC	1660
Db	513	ValAlaIleAsnAlaSerSerGluLeuLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsn	532
Qy	1661	AACACAGAAATCTTCCTTGGAGTGATAAGACAATGTGATAGACCAAGATAGAGAAGCAA	1720
Db	533	AsnHisArgIlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGln	552
Qy	1721	GCGAGGATTTAGCATTCCTCGGTGCGGTGCAACAGTTGAGAGCTCATCAAAACCAG	1780
Db	553	AlaLysAspLeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGln	572
Qy	1781	AAGGAATCTCATTGTTGTAGTGCT	1804
Db	573	ArgGluSerHisPheValSerAla	580
RESULT 4			
ID	Q6PSU4	PRELIMINARY;	PRT; 428 AA.
AC	Q6PSU4		
DT	05-JUN-2004	(TrEMBLrel. 27, Created)	
DT	05-JUN-2004	(TrEMBLrel. 27, Last sequence update)	
DT	05-JUN-2004	(TrEMBLrel. 27, Last annotation update)	
DE	Conararchin (Fragment).		

OS	Arachis hypogaea (Peanut).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC	eurosid I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;
OC	Arachis.
CX	NCBI_TaxID=3818;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Wang L., Liao B., Li H., Yan Y., Lin X., Huang S.;
RL	Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AY581851; AAT00596.1; -.
DR	GO: GO:0045735; F:nutrient reservoir activity; IEA.
DR	InterPro; IPR006045; Cupin_region.
DR	InterPro; IPR007113; Cupin_region.
DR	InterPro; IPR011051; Rn1C_like_cupin.
PF	Pfam; PF00190; Cupin; 2.
FT	NON TER
FT	1
SQ	SEQUENCE 428 AA; 48094 MW; FIA32DAEC3AD06F CRC64;
Alignment Scores:	
Pred. No.:	3.02e-134 Length: 428
Score:	2172.50 Matches: 427
Percent Similarity:	99.77% Conservative: 0
Best Local Similarity:	99.77% Mismatches: 0
Query Match:	60.99% Indels: 1
DB:	2 Gaps: 1
US-10-728-051-1 (1-2032) x Q6PSU4 (1-428)	
QY	644 TTTCCACCAAGGTCACAGCAGTTTCACAAATCTCCAGAAATCACCGTATTGTGCAGATCGAG 703
DB	2 PheaspGlnargSerArgGlnPheGlnAsnLeuGlnAsnHisargIleValGlnIleGlu 21
QY	704 GCCAAACCTAACACTCTTGTCTTCTTCCCAAGCAGCTGATCGTGATAACATCTCTTGTTC 763
DB	22 AlalysProAsnThrLeuValLeuProLysHisAlaAspAlaAspAsnIleLeuValIle 41
QY	764 CAGCAAGGGCAAGCCAGCTGACCGTAGCAATGGCAATACAGAAAGAGCTTTAACTTT 823
DB	42 GlnGlnGlyGlnAlaThrValThrValAlaAsnGlyAsnAsnArgLysSerPheAsnLeu 61
QY	824 GACGAGGGCCATGCACCTCAGAATCCCATCCGTTTTCATTTCCYACATCTTTGAACCCGCAT 883
DB	62 AspGluGlyHisAlaLeuArgIleProSerGlyPheIleSerTyrlleLeuAsnArgHis 81
QY	884 GACAAACAGAACTCAGAGTAGCTAAATCTCCATGCCGTTAACACACCCGCCAGTTT 943
DB	82 AspAsnGlnAsnLeuArgValAlaLysIleSerMetProValAsnThrProGlyGlnPhe 101
QY	944 GAGATTCTTCCGGCGAGCCGAGACCAATCATCTACTTGCAGGGCTTCAGCAGG 1003
DB	102 GluaspPheProAlaSerSerArgAspGlnSerSerTyrlleGlnGlyPheSerArg 121
QY	1004 AATACGTTGAGCGCGCTTCAATGCCGAATTCAAATCAGATACGAGGGTGCTCTTAGAA 1063
DB	122 AsnThrLeuGluAlaAlaPheAsnAlaGluPheAsnGluIleArgArgValLeuLeuGlu 141
QY	1064 GAGAATCGAGAGGTGACGAAGAGAGAGAGGCGAGAGGCGATGGAGTACTCGAGTAGT 1123
DB	142 GluAsnAlaGlyGlyGluGlnGluGluArgGlyGlnArgArgTrpSerThrArgSerSer 161
QY	1124 GAGACAAATGAGGAGTGATAGTCNAAGTGTCAAAGGAGCAGCTTGGAAGACTTACTTAAG 1183
DB	162 GluAsnAsnGluGlyValIleValLysValSerLysGluHisValGluGluLeuThrLys 181
QY	1184 CACGCTAAATCCGCTCAAAGAAAGGCTCCGAAGAAGAGGAGATATCACCAACCCCAATC 1243
DB	182 HisAlalysSerValSerLysLysGlySer---GluGluGlyAspIleThrAsnProIle 200
QY	1244 AACTTGAGAGAGCGGAGCCCGATCTTTCTTAAACAACTTTGGGAAGTTATTGTAGGTGAAG 1303
DB	201 AsnLeuArgGluGlyGluProAspLeuSerAsnAsnPheGlyLysLeuPheGluValLys 220

```
QY 1304 CCAGACAAGAAAGAACCCAGCTTCAGGACCTGGACATGATGCTCACCTGTGTAGAGATC 1363
Db 221 ProAspLysAsnProGlnLeuGlnAspLeuAspMetMetLeuThrCysValGluIle 240
QY 1364 AAAGAAGGAGCTTTGATGCTCCACACTTCAACTCAAGGCCATGTTATCGTCGTC 1423
Db 241 LysGluGlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValVal 260
QY 1424 AACAAAGGAGCTGAAACCTTGAACCTCGCTGCTTAAGAAAGAGCAACACAGAGGGGA 1483
Db 261 AsnLysGlyThrGlyAsnLeuGluLeuValAlaValArgLysGluGlnGlnArgGly 280
QY 1484 CGCGCGGAGAAAGAGGAGGACGAGAGACGAGAGAGAGGAGGAGTAAACAGAGAGTGCCT 1543
Db 281 ArgArgGluGluGluAspGluAspGluGluGluGlySerAsnArgGluValArg 300
QY 1544 AGGTACACGCGAGGTTGAAGGAGGAGGATGTTTATCATGTCAGCAGCTCATCCAGTA 1603
Db 301 ArgTyrThrAlaArgLeuLysGluGlyAspValPheIleMetProAlaAlaHisProVal 320
QY 1604 GCCATCAACGCTTCCCTCCGAACTCCATCTGCTTGGCTTCGGTATCAACGCTGAAACAAAC 1663
Db 321 AlaIleAsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsnAsn 340
QY 1664 CACAGATCTTCCCTTCAGGCTGATAAGGACAAATGTGATGACAGACATAGAGAACCAAGCG 1723
Db 341 HisArgIlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGlnAla 360
QY 1724 AAGATTAGCATTCCTCGGTGCGGTGAACAAGTGTGAGAGCTCATCAAAACCGAGAG 1783
Db 361 LysAspLeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGlnLys 380
QY 1784 GAATCTCACTTGTGAGTGTGCTGCTCAATCTCAATCTCAATCTCAATCTCGTCTCGTCTGAG 1843
Db 381 GluSerHisPheValSerAlaArgProGlnSerGlnSerGlnSerProSerSerProGlu 400
QY 1844 AAAGAGTCTCCTGAGAAGAGGATCAAGAGAGAGAAACCAAGAGGAGGAGGTCCTCCTC 1903
Db 401 LysGluSerProGluLysGluAspGlnGluGluAsnGlnGlyGlyLysGlyProLeu 420
QY 1904 CTTTCAATTTTGAAGGCTTTTAAC 1927
Db 421 LeuSerIleLeuLysAlaPheAsn 428
RESULT 5
Q6PSUS PRELIMINARY; PRT; 299 AA.
ID Q6PSUS AC Q6PSUS;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Conarachin (Fragment).
OS Arachis hypogaea (Peanut).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;
OC Arachis.
OX NCBI_TaxID=3818;
RN [1]_TaxID=3818;
RP SEQUENCE FROM N.A.
RA Wang L., Liao B., Li H., Yan Y., Lin X., Huang S.;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY581850; AAT00595.1; -.
DR GO; GO:0045735; F:nutrient reservoir activity; IEA.
DR InterPro; IPR006045; Cupin.
DR InterPro; IPR007113; Cupin_region.
DR InterPro; IPR01051; RmlC_like_cupin.
DR Pfam; PF00190; Cupin; 1.
FT NON_TER 1
SQ SEQUENCE 299 AA; 33604 MW; 594C3AB0C2FD49EB CRC64;
```

Alignment Scores:

Pred. No.: 8.28e-90 Length: 299

RESULT 6

```
Score: 1495.50 Matches: 296
Percent Similarity: 97.39% Conservative: 2
Best Local Similarity: 96.73% Mismatches: 1
Query Match: 41.98% Indels: 7
DB: 2 Gaps: 1
US-10-728-051-1 (1-2032) x Q6PSU5 (1-299)
QY 1010 TTGGAGGCGCGCTTCAATGCGGAATTCATGAGATACGAGGGTCTGTTTGAAGAGAAAT 1069
Db 1 LeuGluAlaAlaPheAsnAlaGluPheAsnGluIleArgArgValLeuLeuGluAsn 20
QY 1070 GCAGAGGTGAGCAAGAGAGAGAGGAGGCGAGGCGAGTACTCGGAGTAGTGAGAAC 1129
Db 21 AlaGlyGlyGluGlnGluArgGlyGlnArgArgTrpSerThrArgSerSerGluAsn 40
QY 1130 AATGAAGGAGGTATAGTCAAAAGTGTCAAAGAGACAGCTTGAAGAACTTACTAAGCACGCT 1189
Db 41 AsnGluGlyValIleValLysValSerLysGluHisValGluGluLeuThrLysHisAla 60
QY 1190 AAATCCGTCTCAAGAAAGGCTCCGAAGAAGAGGAGATATCAACCAATCAACTTGTG 1249
Db 61 LysSerValSerLysLysGlySerGluGluGluGlyAspIleThrAsnProIleAsnLeu 80
QY 1250 AGAGAAGGCGAGCCCGATCTTCTAACTTTGGGAAGTTATTTGAGGTGAAGCCAGAC 1309
Db 81 ArgGluGlyGluProAspLeuSerAsnAsnPheGlyLysLeuPheGluValLysProAsp 100
QY 1310 AAGAAGAACCCCGAGCTTCAGGACCTGACATGATGCTCACCTGTGTAGAGATCAAGAA 1369
Db 101 LysLysAsnProGlnLeuGlnAspLeuAspMetMetLeuThrCysValGluIleLysGlu 120
QY 1370 GGAGCTTTGATGCTCCACACTTCAACTCAAGGCCATGCTTATCGTCTGCTCAACAA 1429
Db 121 GlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValValValAsnLys 140
QY 1430 GCACTCGGAAACCTTGAACCTCGTGGCTGTAAGAAAGAGCAACAACAGAGGGGACGCGG 1489
Db 141 GlyThrGlyAsnLeuLeuValAlaValArgLysGluGlnGlnGlnArgGlyArgArg 160
QY 1490 GAAGAAGAGGAGGAGCAAGACCAAGAAAGAGGAGGAAAGTAAACAGAGAGTGCCTAGGTAC 1549
Db 161 GluGluGluLysAspGluAspGlnGluGluGlySerAsnArgGluValArgArgTyr 180
QY 1550 ACAGCGAGTTGAAGAGGCGATGTTTCATCATGTCAGCAGCTCATCCAGTAGCCATC 1609
Db 181 ThrAlaArgLeuLysGluGlyAspValPheIleMetProAlaAlaHisProValAlaIle 200
QY 1610 AACGCTTCTCCGAACCTCCATCTGCTTGGCTTCGGTATCAACGCTGAAACCAACCAAGA 1669
Db 201 AsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsnAsnHisArg 220
QY 1670 ATCTTCTTTCAGGTGATAGGACAAATGTGATAGACCATAGAGAGCAAGCAGGAGGAT 1729
Db 221 IlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGlnAlaLysAsp 240
QY 1730 TTAGCATTCCTCGGTGCGGTGAACTTGAAGCTCATCAAAACCAAGAGGAGGAAATCT 1789
Db 241 LeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGlnArgGluSer 260
QY 1790 CACTTTGTGAGTCTCGTCTCAATCTCAATCTCAATCTCCGCTCTCTCTGAGAAAGAG 1849
Db 261 HisPheValSerAlaArgProGlnSerGlnSerProSer----- 273
QY 1850 TCTCTTGAGAAAGAGGATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1909
Db 274 SerProGluLysGluAspGlnGluGluAsnGlnGlyGlyLysGlyProLeuLeuSer 293
QY 1910 ATTTTGAAGGCTTTTAAC 1927
Db 294 IleLeuLysAlaPheAsn 299
```



```
Q6PSU6
ID Q6PSU6 PRELIMINARY; PRT; 303 AA.
AC Q6PSU6;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Conarachin (Fragment).
OS Arachis hypogaea (Peanut).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid1; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;
OC Arachis.
OC NCBI_TaxID=3818;
OX [1]
RN SEQUENCE FROM N.A.
RP Wang L., Liao B., Li H., Yan Y., Lin X., Huang S.;
RA Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
RL EMBL; AY581849; AAT00594.1; -.
DR GO; GO:0045735; F.nutrient reservoir activity; IEA.
DR InterPro; IPR006045; Cupin.
DR InterPro; IPR007113; Cupin region.
DR InterPro; IPR011051; RmlC_like_cupin.
DR Pfam; PF00190; Cupin; 1.
FT NON TER 1
SQ SEQUENCE 303 AA; 34133 MW; 5B4E21BEB48654EC CRC64;

Alignment Scores:
Pred. No.: 7.84e-86 Length: 303
Score: 1435.00 Matches: 291
Percent Similarity: 95.50% Conservative: 6
Best Local Similarity: 93.57% Mismatches: 2
Query Match: 40.29% Indels: 12
DB: 2 Gaps: 4

US-10-728-051-1 (1-2032) x Q6PSU6 (1-303)
QY 1004 AATACGTGAGCGCGCTTCAATGCGAATTCAATGAGATACGAGGCGTGTAGAA 1063
Db 2 AsnThrLeuGluAlaPheAsnAlaGluPheAsnGluLeuArgValLeuLeuGlu 21
QY 1064 GAGAAATCAGAGGTGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 1123
Db 22 GluAsnAlaGlyGluGluGluGluGluGluGluGluGluGluGluGluGlu 41
QY 1124 GAGAAATCAGAGGTGAGTCAAGTGTCAAGGAGGAGGAGGAGGAGGAGGAGG 1183
Db 42 Asp---AsnGluGlyValIleValIleValIleValIleValIleValIleVal 60
QY 1184 CAGCTAAATCGTCTCAAGAAAGGCTCCGAGAGAGGAGGAGGAGGAGGAGGAG 1243
Db 61 HisAlaLysSerValSerLysLysGlySerGluGluGluGluGluGluGluGlu 79
QY 1244 AACTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1303
Db 80 AsnLeuArgAspGlyGluProAspLeuSerAsnAsnPheGlyArgLeuPheGluVal 99
QY 1304 CCAGACAAGAAAGAACCCCGAGCTTCAAGACCTGAGCATGATGCTACCTGTGTAG 1363
Db 100 ProAspLysLysAsnProGlnLeuGlnAspLeuAspMetLeuThrCysValGlu 119
QY 1364 AAGAAGGAGCTTGTATGCTCCACACTTCACTCAAGGCCATGTTATCGTCGTC 1423
Db 120 LysGluGlyAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleValVal 139
QY 1424 AACAAAGGAAGTGAACCTTGAACCTCGTGGCTGTAAAGAAAGAGCAACAGAGGGA 1483
Db 140 AsnLysGlyThrGlyAsnLeuGluLeuValAlaValArgLysGluGlnGlnGlnA 159
QY 1484 CGCGGGAA-----GAAGAGGAGGAGCAAGACGAAAGAGAGGAGGAGGAGTAACA 1534
Db 160 ArgArgGluGlnGluTrpGluGluGluGluGluGluGluGluGluGluGluGlu 179
QY 1535 GAGGTGCTGATAGGTACACAGCGAGGTGAAGAGGCGATGTGTTTCATCATGCCAG 1594
```

```
Db 180 GluValArgArgTyrThrAlaArgLeuLysGluGlyAspValPheIleMetProAlaAla 199
QY 1595 CATCCAGTAGCCATCAACGCTTCTCCGAACTCCATCTGCTGGCTTGGTATCAACGCT 1654
Db 200 HisProValAlaIleAsnAlaSerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAla 219
QY 1655 GAAAAACACACAGAAATCTCTTCAGGTGATGAGACAAATGTATAGACCATAGAG 1714
Db 220 GluAsnAsnHisArgIlePheLeuAlaGlyAspLysAspAsnValIleAspGlnIleGlu 239
QY 1715 AGCAGAGCGAAGGATTAGCATCTCCCTGGTGGGTGGGTGACACAGTTCGAGAGCTCATCA 1774
Db 240 LysGlnAlaLysAspLeuAlaPheProGlySerGlyGluGlnValGluLysLeuIleLys 259
QY 1775 AACCAAGAGGAATCTCACCTTTGTGAGTGTCTCGCTCAATCTCAATCTCAATCTCCGTCG 1834
Db 260 AsnGlnArgGluSerHisPheValSerAlaArgProGlnSerGlnSerProSer----- 277
QY 1835 TCTCCTGAGAAAGAGTCTCTCTGAGAAAGAGGATCAAGAGAGGAGGAGGAGGAGGAG 1894
Db 278 -----SerProGluLysGluAspGlnGluGluGluGluGluGluGluGluGluGlu 292
QY 1895 GGTCCACTCTTCAATTTTGAAGGCTTTTAAAC 1927
Db 293 GlyProLeuLeuSerIleLeuLysAlaPheAsn 303

RESULT 7
Q7XXT2 PRELIMINARY; PRT; 621 AA.
AC Q7XXT2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Prepro beta-conglycinin alpha prime subunit.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC eurosids1; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RA Urade R., Nakatani H., Nakano C.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB113351; BAC78524.1; -.
DR PDB; 1UIK; X-ray; A/B/C=--
DR GO; GO:0045735; F.nutrient reservoir activity; IEA.
DR InterPro; IPR006045; Cupin.
DR InterPro; IPR007113; Cupin region.
DR InterPro; IPR011051; RmlC_like_cupin.
DR Pfam; PF00190; Cupin; 2.
SQ SEQUENCE 621 AA; 72247 MW; 0DF4B6E33737B7F CRC64;

Alignment Scores:
Pred. No.: 5.1e-77 Length: 621
Score: 1302.00 Matches: 286
Percent Similarity: 61.02% Conservative: 121
Best Local Similarity: 42.88% Mismatches: 170
Query Match: 36.55% Indels: 90
DB: 2 Gaps: 19

US-10-728-051-1 (1-2032) x Q7XXT2 (1-621)
QY 50 ATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 109
Db 2 MetArgAlaArgPhePro-----LeuLeuLeuLeuGlyValValPheLeuAlaSerVal 19
QY 110 TTGCAACCGATGCCAAGTCATCCTTACCAAGAGAAACAGAGAACCCCTCGGCCAG 169
Db 20 SerValSerPheGlyIleAla-----TyrTrpGluLys---GlnAsnProSerHisAsn 36
QY 170 AGGTGCTCCAGATTGTCAACAGGAAACCGGATGACTTGAAGCAAAAGGATCGGAGTCT 229
Db 170 AGGTGCTCCAGATTGTCAACAGGAAACCGGATGACTTGAAGCAAAAGGATCGGAGTCT 229
```

```
Db 37 LysCysLeuArgSerCysAsnSerGluLysAspSerTyrArgAsnGlnAlaCysHisAla 56
QY 230 CGCTGCACC-----AAGCTCGAGTATGATCCTCGTTGT-----GTCATGAT 271
Db 57 ArgCysAsnLeuLeuLysValGluGluGluGluGluCysGluGluGlnIleProArg 76
QY 272 CTTCTGA---GGACACTCGCCACCAACCAACACCTTCCCTCCAGGGGAGGGACACGT 328
Db 77 ProArgProGlnHisProGluArgGluArgGlnGlnHisGlyGluLysGluGluAspGlu 96
QY 329 GGCCGCCAACCCGGAGACTACGATGATACCGC---CGTCAACCCCGAAGAGAGAA--- 382
Db 97 GlyGluGlnProArgProPheProPheProArgProArgGlnProArgGlnGluGlu 116
QY 383 -----GAGGCCCGATCG 394
Db 117 HisGluGlnLysGluGluHisGluTrpHisArgLysGluGluLysHisGlyLysGly 136
QY 395 GGA-----CCAGCTGGACCGAGGAGCGTGAA 421
Db 137 SerGluGluGlnAspGluArgGluHisProArgProHisGlnProHisGlnLysGlu 156
QY 422 AGAGAGAA---GACTGGAGACAACCAAGAGAGATTGGAGCGCAACAGTCAATCAGCAG 478
Db 157 GluGluLysHisGluTrpGlnHisLysGlnGluLysHisGlnGlyLysGluSerGluGlu 176
QY 479 CCAGCGAAATAAGGCCGAGGAGGAGAGAGAGAAAGAG---TGGGGAACACCCAGT 535
Db 177 GluGluLysGlnAspGluAspGluGluGlnAspLysGluSerGlnGluSerGluGlu 196
QY 536 AGCCATGTGTGAGGAAGAAACATCTCGAAC-----AACCTTTCTACTTCCGTC 586
Db 197 SerGluSerGlnArgGluProArgArgHisLysAsnLysAsnProPheHisPheAsnSer 216
QY 587 AGCGGTTTAGCCCTAGCGGAGAACCAACAAACGGTAGGATCCGGCTCGTGCAGAGGTTT 646
Db 217 LysArgPheGlnThrLeuPheLysAsnGlnTyrGlyHisValArgValLeuGlnArgPhe 236
QY 647 GACCAAGGTCAAGGCAAGTTTCAGATCTCCAGATCCACGATTCGAGATCGAGGCC 706
Db 237 AsnLysArgSerGlnGlnLeuGlnAsnLeuArgAspTyrArgIleLeuGluPheAsnSer 256
QY 707 AACCTTAACACTTTGTTTCCCAAGACAGCTGATGCTGATTAACATCTCTGTTATCCAG 766
Db 257 LysProAsnThrLeuLeuLeuProHisHisAlaAspAlaAspTyrLeuIleValIleLeu 276
QY 767 CAAGGCAAGCCCGCTGAGCGTAGCAATGGCAATACAGAGAGCTTTAATCTTGAC 826
Db 277 AsnGlyThrAlaIleLeuThrLeuValAsnAsnAspAspArgAspSerTyrAsnLeuGln 296
QY 827 GAGGCGCATGCATCAGATCCCATCCGCTTTCATTTCTACATCTTGAAACCGCCATGAC 886
Db 297 SerGlyAspAlaLeuArgValProAlaGlyThrThrTyrValValAsnProAspAsn 316
QY 887 AACCGAAGCTCAGTAGTAGTAAATCTCCATGCTCCGCTTAACACACCCGCGCAGTTTGAG 946
Db 317 AspGluAsnLeuArgMetIleThrLeuAlaIleProValAsnLysProGlyArgPheGlu 336
QY 947 GATTTCTTCCCGGACGACCGAGACCAATCATCTTCTACTTTCAGGCGCTTCAGCAGGAAT 1006
Db 337 SerPhePheLeuSerThrGlnAlaGlnGlnSerTyrLeuGlnGlnGlyPheSerLysAsn 356
QY 1007 ACCTTGGAGGCGGCTTCAATCGGGAATCAATGAGATACGGAGGGTCTGTTAGAGAG 1066
Db 357 IleLeuGluAlaSerTyrAspThrLysPheGluGluIleAsnLysValLeuPheGlyArg 376
QY 1067 AATCAGAGGTAGCAA---GAGGACAGGGGAGCGGATGAGTACTCGGAGTAGT 1123
Db 377 GluGluGlyGlnGlnGlnGlyGluGluArgLeuGln----- 388
QY 1124 GAGAACAAATGAGGAGTAGTAGTCAAGAGTCAAGGAGCAGCGTTGAAGAACTTACTAAG 1183
Db 389 -----GluSerValIleValGluIleSerLysGlnIleArgGluLeuSerLys 405
```

```
QY 1184 CACGCTAAATCCGTCTCAAAGAAAGCTCCGAAGAGAGGAGATATACCAACCCCAATC 1243
Db 406 HisAlaLysSerSerSerArgLysThrIleSerSerGlu-----AspLysProPhe 422
QY 1244 AACTTGGAGAGAGGCGAGCCGATCTTTCTAACAACTTTGGAGAGTTATTTGAGTGAG 1303
Db 423 AsnLeuArgSerArgAspProIleTyrSerAsnLysLeuGlyLysLeuPheGluIleThr 442
QY 1304 CCAGACAGAGAGAACCCAGCTTCAGCACTTCAGCACTGGACATGATCCTCACCTGTGTAGATC 1363
Db 443 ProGlu---LysAsnProGlnLeuArgAspLeuAspValPheLeuSerValValAspMet 461
QY 1364 AAAGAGAGAGCTTTGATGCTCCCACTTCAACTCAAAAGGCCATGTTATCGTCGTCGTC 1423
Db 462 AsnGluGlyAlaLeuPheLeuProHisPheAsnSerLysAlaIleValValLeuValIle 481
QY 1424 AACAAAGGAACTGGAAACCTTGAACTCTGCTGGCTGTGTAAGAAAGACCAACAGAGGGA 1483
Db 482 AsnGluGlyGluAlaAsnIleGluLeuValGlyIle---LysGluGlnGlnArgGln 500
QY 1484 CGGCGGAGAGAGAGGAGGAGCAAGACGAGAGAGGAGGAAGTAACAGAGAGGTGGCT 1543
Db 501 GlnGlnGluGluGlnPro-----LeuGluValArg 510
QY 1544 AGGTACACAGCGAGGTTGAAGAGCGGATGTTCATCATGCCAGAGCTCATCCAGTA 1603
Db 511 LysTyrArgAlaGluLeuSerGluGlnAspIlePheValIleProAlaGlyTyrProVal 530
QY 1604 GCATCAACGCTTCTCCGAACTCCATCTGTTGCTTGGCTATCAACGCTGAACCAAC 1663
Db 531 ValValAsnAlaThrSerAspLeuAsnPhePheAlaPheGlyIleAsnAlaGluAsnAsn 550
QY 1664 CACAGATCTTCTGTCAGGTGATAGGACAAATGTGATAGACAGATAGAGAGCAAGCG 1723
Db 551 GlnArgAsnPheLeuAlaGlySerLysAspAsnValIleSerGlnIleProSerGlnVal 570
QY 1724 AAGGATTTAGCATTCCTCGGTCGGGTGAACAAAGTTGAGAGCTCATCAAAACAGAG 1783
Db 571 GlnGluLeuAlaPheProGlySerAlaLysAspIleGluAsnLeuIleLysSerGlnSer 590
QY 1784 GAATCTCACTTGTGAGTCTGCTCTCAATCTCAATCTCAATCTCTCGTCTCTCTGAG 1843
Db 591 GluSerTyrPheValAspAlaGlnProGln----- 600
QY 1844 AAAGAGTCTCTCGAGAAAGAGGATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1903
Db 601 -----GlnLysGluGluGlyAsnLysGlyArgLysGlyProLeu 613
QY 1904 CTTTCAATTTGAAGCTTTT 1924
Db 614 SerSerIleLeuArgAlaPhe 620
```

RESULT 8

```
Q6EBCL PRELIMINARY; PRT; 533 AA.
ID Q6EBCL;
AC Q6EBCL;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE Beta-conglutin.
OS Lupinus albus (white lupine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Genisteae; Lupinus.
OX NCBI_TaxID=3870;
RN [1]
RP SEQUENCE FROM N.A.
RA Monteiro S.A., Freitas R.M., Teixeira A.N., Ferreira R.B.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY500372; AAS97865.1;
DR GO; GO:0045735; F:nutrient reservoir activity; IEA.
DR InterPro; IPR006045; Cupin.
```


DE Beta-conglycinin alpha prime subunit.
OS Glycine max (Soybean)
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RA Fukazawa C.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB030838; BAB64303.1; -;
DR HSSP; P25974; 1IPJ.
DR GO; GO:0045735; F:nutrient reservoir activity; IEA.
DR InterPro; IPR006045; Cupin.
DR InterPro; IPR007113; Cupin region.
DR InterPro; IPR01051; Rmlc_Like_cupin.
DR Pfam; PF00190; Cupin; 2.
SQ SEQUENCE 621 AA; 72138 MW; 0196CAD3C6B566DC CRC64;

Alignment Scores:
Pred. No.: 7,76e-76 Length: 621
Score: 1284.00 Matches: 284
Percent Similarity: 59.56% Conservative: 118
Best Local Similarity: 42.07% Mismatches: 167
Query Match: 36.05% Indels: 106
DB: 2 Gaps: 19

US-10-728-051-1 (1-2032) x Q948Y0 (1-621)

QY 50 ATGAGAGGGAGGTTCTCCAGTATGCTCTTCTAGGATCCTTGTCTGGTTCAGTT 109
DB 2 MetArgAlaArgPhePro-----LeuLeuLeuLeuGlyValValPheLeuAlaSerVal 19

QY 110 TCTGCAACGATGCCAGTATCATCTTACAGAGAAACACAGAACCCCTCGGCCAG 169
DB 20 SerValSerPheGlyIleAla-----TyrTrpGluLys---GlnAsnProSerHisAsn 36

QY 170 AGGTGCTCCAGAGTCTCAACAGGACCGGATGACTTGAAGCAAAAGGATCGAGTCT 229
DB 37 LysCysLeuArgSerCysAsnSerGluLysAspSerTyrArgAsnGlnAlaCysHisAla 56

QY 230 CGCTGCACC-----AAGCTCGATGATGATCCTCGTTGT-----GTCATGAT 271
DB 57 ArgCysAsnLeuLeuLysValGluGluGluGluCysGluGluGluGlnIleProArg 76

QY 272 CCTCGA---GGACACTGGCACCACCAACCAACGTTCCCTCCAGGGAGCGACACGT 328
DB 77 ProArgProGlnHisProGluArgGluArgGlnGlnHisGlyGluLysGluLysGlu 96

QY 329 GGCCGCCAACCCGGAGACTACGATGATGACCGC---CGTCAACCCCGAAGAGAGGA 385
DB 97 GlyGluGlnProArgProPheProArgProArgProArgGlnProArgGlnGlu---- 114

QY 386 GGCCGATGGGACACAGCTGGACCGAGGAGCGTGAAGAGAGAGAGACTGGACACCA 445
DB 115 -----GlyGluHisGluGlnLysGluGluHisGluTrpHisArgLys 128

QY 446 AGAGAA-----GATTGGAGGGGACCAAGTATCATGACGACCCAGG 484
DB 129 GluGluLysHisGlyGlyLysGlySerGluGluGluGlnAspGlyArgGluHisProArg 148

QY 485 AAAATAAGGCC----- 496
DB 149 ProHisGlnProHisGlnLysGluGluGluLysHisGluTrpGlnHisLysGlnLys 168

QY 497 ---GAAGAGAGAGAGAGACAGAG----- 520
DB 169 HisGlnGlyLysGluSerGluGluGluGluGluAspGlnAspGluAspGluGluGlnAsp 188

QY 521 -----TGGGAACACACAGGTAGCCATGTGAGGGAAGAAACATCTCGGAAC----- 565
DB 189 LysGluSerGlnGluSerGluGlySerGluSerGluGluGluGluProArgGlyHisLysAsn 208

QY 566 ---AACCCCTTCTACTCCCGTCAAGCGGTTTACACCCCGCTACGGGAACCAAAACCGT 622
DB 209 LysAsnProPheHisPheAsnSerLysArgPheGlnThrLeuPheLysAsnGlnTyrGly 228

QY 623 AGGATCCGGCTCTGCAGAGTTTGCACCAAGAGTCAAGGCGAGTTTCAGAAATCCAGAA 682
DB 229 HisValArgValLeuGlnArgPheAsnLysArgSerGlnGlnLeuGlnAsnLeuArgAsp 248

QY 683 CACCGTATTGTGCAGATCGAGGCCAAACCTAATCACTCTTGTCTTCTCCCAAGCACGCTGAT 742
DB 249 TyrArgGlnLeuGluPheAsnSerLysProAsnThrLeuLeuLeuProHisHisAlaAsp 268

QY 743 GCTGATAACATCTTGTATTATCCAGCAGGGCAAGCCACCGTACCGTACGTAAGCAATCGCAAT 802
DB 269 AlaAspTyrLeuIleValIleLeuAsnGlyThrAlaIleLeuThrLeuValAsnAsnAsp 288

QY 803 AACAGAAAGAGCTTTAATCTTCACGAGGCCCATGCATCCAGATCCCATCCGCTTTCATT 862
DB 289 AspArgAspSerTyrAsnLeuGlnSerGlyAspAlaLeuArgValProAlaGlyThrThr 308

QY 863 TCCTATCATCTTGAACCGCCATGCAACACCAAGAACCTTCAGATGAGTAAATCTCCATGCC 922
DB 309 TyrTyrValValAsnProAspAsnAspGluAsnLeuArgMetIleThrLeuAlaIlePro 328

QY 923 GTTAAACACACCGCGCGATTTGAGGATTTCTTCGCGCGAGCAGCGAGCAACATCATCC 982
DB 329 ValAsnLysProGlyArgPheGluSerPhePheLeuSerSerThrGlnAlaGlnGlnSer 348

QY 983 TACTTGCAGGCTTCAGCAGGAATACGTTGGAGCGCGCTTCAATCGCGAATTCATAG 1042
DB 349 TyrLeuGlnGlyPheSerLysAsnIleLeuGluAlaSerTyrAspThrLysPheGluGlu 368

QY 1043 ATACGAGCGGCTCTGTAGAGAGAAATGAGGAGGTGAGCAA---GAGGAGAGAGGGCAG 1099
DB 369 IleAsnLysValLeuPheGlyArgGluGluGlyGlnGlnGlnGlyGluArgLeuGln 388

QY 1100 AGCGATGAGTACTCGGAGTAGTGAGAACATGAGGAGTGATAGTCAAAAGTCAAG 1159
DB 389 -----GluSerValIleValGluLysSerLys 397

QY 1160 GAGCAGCTTGAAGAACTTACTAAGCAGCTAAATCCGTCAAAGAAAGGCTCCGAAGAA 1219
DB 398 LysGlnIleArgGluLeuSerLysArgAlaLysSerSerSerArgLysThrIleSerSer 417

QY 1220 GAGGAGATATACCAACCAATCACTTGAGAGAGGGCAGCGCATCTTCTTCAACAC 1279
DB 418 Glu-----AspLysProPheAsnLeuArgSerArgAspProIleTyrSerAsnLys 434

QY 1280 TTTGGGAAGTATTTCAGCTGAAGCGTGAAGCAGACAGAGAACCCCGCTTCAGGACCTGGAC 1339
DB 435 LeuGlyLysLeuPheGluIleThrProGlu---LysAsnProGlnLeuArgAspLeuAsp 453

QY 1340 ATGATCTCACCTGTGTAGAGATCAAGAGAGAGCTTTGATGCTCCACACTTCAACTCA 1399
DB 454 ValPheLeuSerValValAspMetAsnGluGlyAlaLeuPheLeuProHisPheAsnSer 473

QY 1400 AAGGCATGTTATCGTCTGTCACAAAGAGAACTGGAAACCTTGAACCTCGTGGCTGTA 1459
DB 474 LysAlaIleValValLeuValIleAsnGluGlyGluAlaAsnIleGluLeuValGlyIle 493

QY 1460 AGAAAGAGCAACACAGAGGGGAGCGCGGAGAGAGAGGAGGACGACGAAGAAAGAG 1519
DB 494 ---LysGluGlnGlnGlnArgGlnGlnGlnGluGlnPro----- 506

QY 1520 GAGGGAAGTAACAGAGAGGTGCGTAGGTATACAGCGAGGTGTAAGGAAGCGCATGTGTC 1579
DB 507 -----LeuGluValArgLysTyrArgAlaGluLeuSerGluGlnAspIlePhe 522

QY 1580 ATCATCGCAGCAGCTCATCAGTAGCCATCAACGCTTCCTCCGAACCTCCATCTGCTTGGC 1639
DB 523 ValIleProAlaGlyTyrProValValValAsnAlaThrSerAspLeuAsnPhePheAla 542

QY 1640 TTCGGTATCAACGCTCAAAACACCAACCAAGATCTTCTCTTCAGGCTGTAAGGACAATGTG 1699

```
Db 543 PheGlyIleAsnAlaGluAsnAnGlnArgAsnPheLeuAlaGlySerLysAspAsnVal 562
QY 1700 ATAGACCAGATAGAGAAGCAAGGATTTAGCATTCCTCGGTGGGTGAACAAGTT 1759
Db 563 IleSerGlnIleProSerGlnValGlnGluLeuAlaPheLeuGlySerAlaLysAspIle 582
QY 1760 GAGAGCTCATCAAAAACAGAGGAATCTCATTTGTGTAGTGTCTCTCAATCTCAA 1819
Db 583 GlnAsnLeuIleLysSerGlnSerGluSerTyrPheValAspAlaGlnProGln----- 600
QY 1820 TCTCAATCTCCGTCTCTCTGAGAAAGAGTCTCTCTGAGAAAGAGGATCAAGAGGAGAA 1879
Db 601 -----GlnLysGluGluGly 605
QY 1880 AACCAAGGAGGAGGAGGTCACCTCTTCAATTTGAAGGCTTTT 1924
Db 606 AsnLysGlyArgLysGlyProLeuSerSerIleLeuArgAlaPhe 620

RESULT 10
Q94LX2 PRELIMINARY; PRT; 605 AA.
AC Q94LX2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Beta-conglycinin alpha subunit.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosids;
OC eucotids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21327318; PubMed=11434464;
RA Yoshino M., Kanazawa A., Tsutsumi K., Nakamura I., Shimamoto Y.;
RT "Structure and characterization of the gene encoding alpha subunit of
RT soybean beta-conglycinin."
RL Genes Genet. Syst. 76:99-105(2001).
DR EMBL; AB051865; BAB56161.1; -.
DR PIR; S20007; S20007.
DR HSSP; P25974; 1IPJ.
DR GO; GO:0045735; F:nutrient reservoir activity; IEA.
DR InterPro; IPR006045; Cupin.
DR InterPro; IPR007113; Cupin region.
DR InterPro; IPR011051; RmlC_like_cupin.
DR Pfam; PF00190; Cupin; 2.
SQ SEQUENCE 605 AA; 70306 MW; 8ACB6F8532662984 CRC64;

Alignment Scores:
Pred. No.: 1.04e-75 Length: 605
Score: 1282.00 Matches: 285
Percent Similarity: 59.37% Conservative: 108
Best Local Similarity: 43.05% Mismatches: 173
Query Match: 35.99% Indels: 96
DB: 2 Gaps: 18

US-10-728-051-1 (1-2032) x Q94LX2 (1-605)
QY 50 ATGAGAGGAGGGTTTCTCCACTGATGCTGCTAGGATCTTGTCTCGGCTTCAGTT 109
Db 2 MetArgAlaArgPhePro-----LeuLeuLeuGlyLeuValPheLeuAlaSerVal 19
QY 110 TCTGCAACGATGCCAAGTCATCATCTTACAGAGAAACACAGAACCCCTCGGCCAG 169
Db 20 SerValSerPheGlyIleAla-----TyrTrpGluLys---GluAsnProLysHisAsn 36
QY 170 AGTGCTCCAGAGTTGTCAACAGGAACCGGATGACTTGAAGCAAAAGGATCGAGTCT 229
Db 37 LysCysLeuGlnSerCysAsnSerGluArgAspSerTyrArgAsnGlnAlaCysHisAla 56
QY 230 CGCTGCAACCAAGCTCGAGTATGAT-----CCTCGTTGT 262
|||||
```

```
Db 57 ArgCysAsnLeuLeuLysValGluLysGluGluCysGluGluGlyGluIleProArg--- 75
QY 263 GTCTATGATCTCGAGACACACTGGCACCAACCAACAGTTTCCCTCTCAGGGAGCGG 322
Db 76 -----ProArgProArg-----ProGlnHisProGlu 84
QY 323 ACAGTGGCGCCCAACCCGAGACTACGATGATGAC----- 358
Db 85 ArgGluProGlnGlnProGlyGluLysGluLysGluAspGluAspGluProIle 104.
QY 359 -----CGCGTCAACCCCGAAGAGAG-----GAAGGAGGCCGA 391
Db 105 ProPheProArgProGlnProArgGlnGluGluHisGluGlnArgGluGluGlnGlu 124
QY 392 TGG-----GCACCACTGGACCCGAGGAGCGGTGAAGAGAGAA 430
Db 125 TrpProArgLysGluGluLysArgGlyGluLysGlySerGluGluGluAspGluAspGlu 144
QY 431 GACTGGAGACACCAAGAGAGAT-----TGAGGGGACCAAGTGTATCAGCAGCCACGG 484
Db 145 AspGluGluGlnAspGluArgGlnPheProPheProArgProHisGlnLysGluGlu 164
QY 485 AAAATAAGGCCCGCAAGAGAGAGAGAGAGAGTGGGGAAACACCAGGTAGCCATGTG 544
Db 165 ArgLysGlnGluGluAspGluAspGluGlnGlnArgGlnSerGluGluSerGluAsp 184
QY 545 AGGGAA-----GAAACATCTCGGAACAACCTTTCTACTTCCCGTCAAGCGGTTTACG 598
Db 185 SerGluLeuArgArgHisLysAsnLysAsnProPheLeuPheGlySerAsnArgPheGlu 204
QY 599 ACCGCTACGGGAACCAAAACGGTAGATCCGGGTCTCGCAGAGGTTTGCACCAAGGTCA 658
Db 205 ThrLeuPheLysAsnGlnTyrGlyArgIleArgValLeuGlnArgPheAsnGlnArgSer 224
QY 659 AGGCAGTTTCAGAAATCTCCAGATCACCGTATTGTGCAGATCGAGGCCAAACCTTAACACT 718
Db 225 ProGlnLeuGlnAsnLeuArgAspTyrArgIleGluGluPheAsnSerLysProAsnThr 244
QY 719 CTTGTTCTTCCCAAGCAGCGTGTATGCTGATAACATCTTGTATCCAGCAAGGCGAAGCC 778
Db 245 LeuLeuLeuProAsnHisAlaAspAlaAspTyrLeuIleValIleLeuAsnGlyThrAla 264
QY 779 ACCGTACCGTAGCAATTCGAATACAGCAAGAGCTTTAATCTTGCAGGGGCCATGCA 838
Db 265 IleLeuSerLeuValAsnAsnAspArgAspSerTyrArgLeuGlnSerGlyAspAla 284
QY 839 CTCAGAAATCCCATCCCGTTCATTTCTTCAACCGCATCTGACCAACCAAGAACCTC 898
Db 285 LeuArgValProSerGlyThrThrTyrValValAsnProAsnAsnGluAsnLeu 304
QY 899 AGAGTAGCTAAATCTCCATGCCCGTTAAACACACCCGCGCAGTTTGAGGATTTCTCCCG 958
Db 305 ArgLeuIleThrLeuAlaIleProValAsnLysProGlyArgPheGluSerPheLeu 324
QY 959 GCGAGACGGGAGACCAATCATCTTCTAGCGGCTTCAGCAGGAATACGTTGGAGGCC 1018
Db 325 SerSerThrGluAlaGlnGlnSerTyrLeuGlnGlyPheSerArgAsnIleLeuGluAla 344
QY 1019 GCCTTCAATCGCAATTCATGAGATACGAGGCGTCTGTTAGAGAGATGCGAGGAGGT 1078
Db 345 SerTyrAspThrLysPheGluGluIleAsnLysValLeuPheSerArgGluGluGly--- 363
QY 1079 GAGCAAGAGAGAGAGGAGCGGATGAGTCTCGGAGTAGTGAGAAACAATGAAGGA 1138
Db 364 -----GlnGlnGlnGlyGlnArgLeu-----GlnGluSer 374
QY 1139 GTGATAGTCAAGGTCAAGGAGAGCGGTTGAAGAACTTACTAAGCAACGCTAATCCGTC 1198
Db 375 ValIleValGluIleSerLysGluGlnIleArgAlaLeuSerLysArgAlaLysSerSer 394
QY 1199 TCAAGAAAGGCTCCGAGAGAGGAGGATATCACCACCAACCTCACTTGTGAGAGAGGC 1258
Db 395 SerArgLysThrIleSerSerGlu-----AspLysProPheAsnLeuArgSerArg 411
|||||
```

```
QY 1259 GAGCCCGATCTTTTCTAACAACCTTTGGGAAGTTATTTGAGGTGAAGCCAGACAGAGAAGAAC 1318
DB 1412 AppProLeyr:SerAsnLysLeuGlyLysPhePheGluLeuThrProGlu---LysAsn 430
QY 1319 CCCAGCTTCAGACCTGACATGATGTCACCTGTGTAGATCAAAAGAGAGCTTTG 1378
DB 1431 ProGlnLeuArgAspLeuAspIlePheLeuSerIleValAspMetAsnGluGlyAlaLeu 450
QY 1379 ATGCTCCACACTCAACTCAAGGCGCATGTTGTCGTCTGCTCAACAAAGGAACTGGA 1438
DB 1451 LeuLeuProHisPheAsnSerLysAlaIleValIleLeuValIleAsnGluGlyAspAla 470
QY 1439 AACCTTGAACCTCGTGTGTAAGAAAGAGCAACACAGAGGCGCGGGAAGAGAG 1498
DB 1471 AsnIleGluLeuValGlyLeu---LysGluGlnGlnGlnGlnGlnGluGln 489
QY 1499 GAGGCGAAGCAAGCAAGAGAGGAGGAAGTAACAGAGAGGTGGTACACAGCGAGG 1558
DB 1490 Pro-----LeuGluValArgLysTyrArgAlaGlu 499
QY 1559 TTGAAGGAGCGCATGTTTCATCGCCAGCAGCTCATCCAGTAGCCATCAACGCTCC 1618
DB 1500 LeuSerGluGlnAspIlePheValIleProAlaGlyTyrProValValValAsnAlaThr 519
QY 1619 TCCGAACCTCCATCTGCTTTCGCTTTCGCTTATCAACGCTGAAACCAACACAGAAATCTTCTT 1678
DB 1520 SerAsnLeuAsnPhePheAlaIleGlyIleAsnAlaGluAsnAsnGlnArgAsnPheLeu 539
QY 1679 GCAGGTGATAAGACAAATGTGTAGACAGATAGAGCAAGCAAGCAAGATTTAGCATTC 1738
DB 1540 AlaGlySerGlnAspAsnValIleSerGlnIleProSerGlnValGlnGluLeuAlaPhe 559
QY 1739 CTGGGTGCGGTCAACAAGTTGAGAGCTCATCAAAAACAGAAAGAACTCTCACTTTGG 1798
DB 1560 ProGlySerAlaGlnAlaValGluLysLeuLeuLysAsnGlnArgGluSerTyrPheVal 579
QY 1799 AGTGCTCGTCTCAATCTCAATCTCAATCTCCGCTCTCTCGAGAAAGAGTCTCTCGAG 1858
DB 1580 AspAlaGlnProLysLys----- 585
QY 1859 AAGAGGATCAAGAGGAGGAAACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1918
DB 1586 -----LysGluGluGlyAsnLysGlyArgLysGlyProLeuSerSerIleLeuArg 602
QY 1919 GCTTTT 1924
DB 1603 AlaPhe 604

RESULT 11
GLCA_SOYBN
ID GLCA_SOYBN STANDARD; PRT; 605 AA.
AC P13916;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Beta-conglycinin, alpha chain precursor.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RC SEQUENCE FROM N.A.
RP TISSUE=Cotyledon;
RX MEDLINE=9135860; PubMed=2103438;
RA Sebastiani F.L., Farrell L.B., Schuler M.A., Beachy R.N.;
RT "Complete sequence of a cDNA of alpha subunit of soybean beta-
conglycinin."
RL Plant Mol. Biol. 15:197-201(1990).
CC -|- FUNCTION: Seed storage protein. Accumulates during seed
development and is hydrolyzed after germination to provide a
carbon and nitrogen source for the developing seedling.
```

```
CC -|- SUBUNIT: The alpha-, alpha-, and beta-subunits associate in
CC various combinations to form trimeric proteins.
CC -|- SUBCELLULAR LOCATION: Embryo axis, and cotyledonary membrane-bound
CC vacuolar protein bodies.
CC -|- SIMILARITY: Belongs to the 7S seed storage protein family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X17698; CAA35691.1; -.
DR PIR; S14681; FWSVBA.
DR HSSP; P25974; IIRJ.
DR InterPro; IPR006045; Cupin.
DR InterPro; IPR007113; Cupin region.
DR InterPro; IPR011031; Multihaem cyt.
DR InterPro; IPR011051; RmlC_like_cupin.
DR Pfam; PF00190; Cupin; 2.
KW Glycoprotein; Multigene family; Seed storage protein; Signal.
FT SIGNAL 1 22
FT PROPEP 23 62
FT CHAIN 63 605 Beta-conglycinin, alpha chain.
FT CARBOHYD 261 261 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 517 517 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 605 AA; 70293 MW; CBEB30506BBBC57 CRC64;

Alignment Scores:
Pred. No.: 1,218-75 Length: 605
Score: 1281.00 Matches: 285
Percent Similarity: 59.21% Conservative: 107
Best Local Similarity: 43.03% Mismatches: 174
Query Match: 35.96% Indels: 96
DB: 1 Gaps: 18

US-10-728-051-1 (1-2032) x GLCA_SOYBN (1-605)
QY 50 ATGAGAGGAGGAGGTTTCTCCACTGATGCTGTGTAGGATCTTGTCTGCTTGTGCTT 109
DB 2 MetArgAlaArgPhePro-----LeuLeuLeuLeuGlyLeuValPheLeuAlaSerVal 19
QY 110 TCTGCAACGATGCCCAAGTCACTTACAGAGAAGAAACAGAGAACCCCTGCGCCAG 169
DB 20 SerValSerPheGlyIleAla-----TyrTrpGluLys---GluAsnProLysHisAsn 36
QY 170 AGTGCTCTCAGATGTTGTCAACAGGACCGGATGACTTGAAGCAAAAGCATGCGAGTCT 229
DB 37 LysCysLeuGlnSerCysAsnSerGluArgAspSerTyrArgAsnGlnAlaCysHisAla 56
QY 230 CGCTGCACCAAGCTCGAGTATGAT-----CTCGTTGT 262
DB 57 ArgCysAsnLeuLeuLysValGluLysGluGluCysGluGluGlyGluLeuProArg--- 75
QY 263 GTCTATGATCTCTGAGGACACACTGSCACCAACCAACACCTTCCCTCCAGGGAGCG 322
DB 76 -----ProArgProArg-----ProGlnHisProGlu 84
QY 323 ACAGTGGCGGCGCAACCGGAGACTACGATGATGAC----- 358
DB 85 ArgGluProGlnGlnProGlyGluLysGluGluAspGluAspGluGlnProArgProIle 104
QY 359 -----CGCGCTCAACCCGAGAGAG-----GAAGGAGGCGCA 391
DB 105 ProPheProArgProGlnProArgGlnGluGluHisGluGlnArgGluGlnGlu 124
QY 392 TGG-----GGACCGCTGGACCGGAGGAGCGTGAAGAGAGAA 430
DB 125 TrpProArgLysGluGluLysArgGlyGlyGlySerGlyGluGluAspGlu 144
QY 431 GACTGAGACACCAACCAAGAGAGAT-----TGGAGGCGACCAAGTATCATCAGCAGCG 484
```

[illegible]

KW Direct protein sequencing; Multigene family; Seed storage protein;

KW	Signal.	1	28	
FT	SIGNAL	29	571	Convicilin.
FT	CHAIN			
SQ	SEQUENCE	571 AA;	66989 MW;	749CFBE2D16D57B CRC64;

Alignment Scores:

Pred. No.:	9,95e-75	Length:	571
Score:	1267.00	Matches:	276
Percent Similarity:	56.67%	Conservative:	98
Best Local Similarity:	41.82%	Mismatches:	146
Query Match:	35.57%	Indels:	140
DB:	1	Gaps:	13

US-10-728-051-1 (1-2032) x CVCA PEA (1-571)

Qy	68	CACTGATGCTGTGCTAGGATACCTTGTCTCGGCTTCAGTTTCTGCAACGCATGCCAAG	127
Db	10	ProLeuLeuPheLeuGlyIleIlePheLeuAlaSerValCysValThrThrAlaAsn	29
Qy	128	TCATCACCTTACCAGAAGAAACAGAGAACCCTCGCGCCACAGAGGTGCCTCCAGAGTTGT	187
Db	29	-----	29
Qy	188	CAACAGAAACCGGATGACTTGAGCAAAAGGCATGCGAGTCTCGTCACCAAGCTCGAG	247
Db	29	-----	29
Qy	248	TATGATCCTCGTTGTGTCTATGATCCTCGAGGACACACTGGCACCAACCAACGATTC	307
Db	30	-----TyrAspGluGlySerGluThrArgVal	38
Qy	308	CCTCCAGGGGAGCGGACACGTCGGCCGCAACCCGGAGACTACGATGATGACCGCTCAA	367
Db	39	---ProGlyGlnArgGluArgGlyArgGlnGluGlyGluGlyArgHis---	56
Qy	368	CCCCGAAGAGAGAAAGGCGCGATGGGGACACAGCTGGACCGAGGAGCGT---	418
Db	57	-----GlyIuTrpArgProSerTyrGluLysGluGluHisGluGlu	70
Qy	418	-----	418
Db	71	GluLysGlnLysTyrArgTyrGlnArgGluLysLysGluGlnLysGluValGlnProGly	90
Qy	419	-----GAAAGAGAAAGACTGGAGACAACCAAGAGAAAGATTGGAG---	460
Db	91	ArgGluArgTrpGluArgGluAspGluGluGlnValGluGluGluTrpArgGlySer	110
Qy	461	-----CGACCAAGTCAATCAGCAGCCACGG	484
Db	111	GlnArgArgGluAspProGluGluArgAlaArgLeuArgHisArgGluGluArgThrLys	130
Qy	485	AAATAAGGCCCGAAGGAGAGAGAGAAACAAGATGGGGAAACCCAGGTAGCCATGTG	544
Db	131	ArgAspArgArgHisGlnArgGluGlyGluGluGlu-----GluArgSer	145
Qy	545	AGGGAAGAAACATCTCGGAACAACCCCTTCTACTTCCGTCGACGCGTTAGCACCCGC	604
Db	146	SerGluSerGlnGluHisArgAsnProPheLeuPheLysSerAsnLysPheLeuThrLeu	165
Qy	605	TACGGGAACCAAAACGGTAGGATCGGGTCTCGAGAGGTTTGACCAAGGTCAAGGCAG	664
Db	166	PheGluAsnGluAsnGlyHisIleArgArgLeuGlnArgPheAspLysArgSerAspLeu	185
Qy	665	TTTCAGAATCTCCAGAATCACCGTATTGTCAGATCGAGCGCAACACCTAACTCTTGT	724
Db	186	PheGluAsnLeuGlnAsnTyrArgLeuValGluTyrArgAlaLysProHisThrIlePhe	205
Qy	725	CTTCCCAAGCAGCTGATGCTGATAACTCTTGTATTCCAGCAAGGCAAGCCACCGTG	784
Db	206	LeuProGlnHisIleAspAlaAspLeuIleLeuValLeuAsnGluLysAlaIleLeu	225
Qy	785	ACGCTAGCAAAATGGCAATAACAGAAGAGCTTTAATCTTGACGAGGCCATGCACCTCAG	844


```
Db 551 GlnLysGluGluGluSerGlnArgLysArgSerProLeuSerSerValLeuAspSerPhe 570
RESULT 13
Q948X9 PRELIMINARY; PRT; 623 AA.
AC Q948X9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Beta-conglycinin alpha-subunit.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RA Fukazawa C.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB030839; BAB64304.1; -.
DR HSSP; P25974; 1IPJ.
DR GO; GO:0045735; F:nutrient reservoir activity; IEA.
DR InterPro; IPR006045; Cupin.
DR InterPro; IPR007113; Cupin region.
DR InterPro; IPR011051; RmC_Like_cupin.
DR Pfam; PF00190; Cupin; 2.
SQ SEQUENCE 623 AA; 72475 MW; EBF033DA1316AE9A CRC64;

Alignment Scores:
Pred. No.: 1-28e-74 Length: 623
Score: 1265.50 Matches: 283
Percent Similarity: 58.22% Conservative: 110
Best Local Similarity: 41.93% Mismatches: 179
Query Match: 35.53% Indels: 103
DB: 2 Gaps: 18

US-10-728-051-1 (1-2032) x Q948X9 (1-623)
QY 50 ATGAGAGGGAGGTTTCTCCAGTATGCTGTGCTAGGATCCTTGTCTGGTTCAGTT 109
Db 1 MetArgAlaArgPhePro-----LeuLeuLeuLeuGlyValPheLeuAlaSerVal 18
QY 110 TGTGCAACGATGCCAAGTCATACCTTACAGAGAGAAACACAGAACCCCTGGCCAG 169
Db 19 SerValSerPheGlyIleAla-----TyrTrpGluLys---GlnAsnProLysHisAsn 35
QY 170 AGTGCTCTCAGAGTCTCAACAGGACCGGATGCTTGAAGCAAAAGGCATCGAGTCT 229
Db 36 LysCysLeuGlnSerCysAsnSerGluArgAspSerTyrArgAsnGlnAlaCysHisAla 55
QY 230 CGTGTGACCC-----AAGCTCGAGTATGATCCTCGTTGCTGTCTATGATCCTCGAGGACAC 283
Db 56 ArgCysAsnLeuLeuLysValGluLysGluGluCys-----GluGluGlyGlu 72
QY 284 ACTGGCAACCAACCAACGTTTCCCTCCAGGGAGCGGACACGTGGCCGCCAACCCGGA 343
Db 73 IleProArgProArgProArgProGlnHisProGluArg---GluProGlnGlnProGly 91
QY 344 GACTACCATGATGACCC-----CGTCAA 367
Db 92 GluLysGluGluAspGluAspGluGlnProArgProIleProPheProArgProArgGln 111
QY 368 CCCGAGAGAGAGAA----- 382
Db 112 ProArgGlnGluGluGluHisGluGlnArgGluGluGlnGluTrpProArgLysGluGlu 131
QY 383 ---GGAGGCGCATGGGA-----CCAGCT 403
Db 132 LysArgGlyGluLysGlySerGluGluGluGlnAspGlyArgGluHisProArgProHis 151
QY 404 GGACCGAGGAGCGTGAAGAGAGAGAGACTGGAGACACCAAGAGAGATTGGAGCGGA 463
Db 152 GlnProHisAspGluAspGluGluGlnAspGluArgGln-----PheProPheProArg 169
```

```
QY 464 CCAAGTCATCAGCAGCCACGGAATAAGG-----CCCGAA 499
Db 170 ProProHisGlnLysGluSerGluGluArgLysGlnGluAspGluAspGluGln 189
QY 500 GGNAGAGAGAGAAACAAGAGTGGGNAACACACAGGTAGCATGTAGGGAAGAAACATCT 559
Db 190 GlnArgGluSerGluGluSerGluSerGluSerGlnArgGluLeuArgArgHisLys 209
QY 560 CGGAACAACACCTTTCTACTTCCCGTCAAGCGGCTTTAGCACCCGCTACGGAACCAAAAC 619
Db 210 AsnLysAsnProPheHisPheGlySerAsnArgPheGluThrLeuPheLysAsnGlnTyr 229
QY 620 GTTAGGATCCGGTCTCTGAGAGGTTTGACAAAGGTCAGAGGCGATTTTCAAGATCTCCAG 679
Db 230 GlyArgIleArgValLeuGlnArgPheAsnGlnArgSerProGlnLeuGlnAsnLeuArg 249
QY 680 AATCACCGTATTGTGAGATCGAGGCGCAACCTTAACACTCTTCTTCTCCNAGCAGCT 739
Db 250 AspTyrArgIleLeuGluPheAsnSerLysProAsnThrLeuLeuLeuProAsnHisAla 269
QY 740 GATGCTGATAAACAATCTTGTATTCCAGCAAGGCGCAAGCCACCGTAGCAAAATGGC 799
Db 270 AspAlaAspTyrLeuLeuAlaIleLeuAsnGlyThrAlaIleLeuSerLeuValAsnAsn 289
QY 800 AATAACAGAAAGAGCTTTAATTTGACGAGGCGCATGCACTCAGATCCCATCCGTTTC 859
Db 290 AspAspArgAspSerTyrArgLeuGlnSerGlyAspAlaLeuArgValProSerGlyThr 309
QY 860 ATTTCCTACATCTTGAACCGCCATGACAAACCAACCTCAGATAGTAGTAATAATCTCCATG 919
Db 310 ThrTyrTyrValValAsnProAspAsnAsnGlnAsnLeuLeuThrLeuAlaIle 329
QY 920 CCGTTTAAACACACCCGCGGATTTGAGGATTTCTTCCGCGAGCAGCAGCCAGCAATCA 979
Db 330 ProValAsnLysProGlyArgPheGluSerPhePheLeuSerSerThrGluAlaGlnGln 349
QY 980 TCTACTTTCAGGCTTTCAGCAAGATACGTCTGGAGCGCCCTTCAATGCGGAATTCAT 1039
Db 350 SerTyrLeuGlnGlyPheSerArgAsnIleLeuGluAlaSerTyrAspThrLysPheGlu 369
QY 1040 GAGATACGGAGGTGCTGTAGNAGAGATGCAAGAGGTGAGCAAGAGGAGAGAGGCGAG 1099
Db 370 GluIleAsnLysValLeuPheSerArgGluGluGly-----GlnGlnGlnGlyGlu 386
QY 1100 AGCGATGGAGTACTCGGAGTAGTGAGAACAAATGAAGGAGTAGTCAAAAGTGTCAAAG 1159
Db 387 GlnArgLeu-----GlnGluSerValIleValGluIleSerLys 399
QY 1160 GAGCAGCTTGAAGAACTTACTAAGCAGCTTAATCCGTCTCAAGAAAGAGCTCCGAAGAA 1219
Db 400 GluGlnIleArgAlaLeuSerLysArgAlaLysSerSerSerArgLysThrIleSerSer 419
QY 1220 GAGGAGATATCACCAACCCATCACTTGAGAGAAGCGCGCCGATCTTTCTTAACAC 1279
Db 420 Glu-----AspLysProPheAsnLeuArgSerArgAspProIleTyrSerAsnLys 436
QY 1280 TTTGGGAAGTTATTTCAGGTGAAGCAGACAAAGAACCCCGAGCTTCAGGACCTGCAC 1339
Db 437 LeuGlyLysPhePheGluIleThrProGlu---LysAsnProGlnLeuArgAspLeuAsp 455
QY 1340 ATGATGCTCACCTGTGTAGAGATCAAGAGAGAGCTTTGATGTCTCCACACTTCAACTCA 1399
Db 456 IlePheLeuSerIleValAspMetAsnGluGlyAlaLeuLeuLeuProHisPheAsnSer 475
QY 1400 AAGGCATCGTTATCGTCTCGTCAACAAGAGAACTGGAAACCTTGAACCTCGGCTGTA 1459
Db 476 LysAlaIleValIleLeuValIleAsnGluGlyAspAlaAsnIleGluLeuValGlyLeu 495
QY 1460 AGAAAAGAGCAACAACAGAGGGGACCGCGGGAAGAGAGGAGGACGGAAGACGAAGAGAG 1519
Db 496 ---LysGluGlnGlnGln-----GluGluGlnGln 504
```



```
Db 312 ThrAsn-----AlaLeuValLysValSerArgGluGlnValGluLeuLysArg 328
Qy 1184 CACCTAAATCCGCTCAAGAAAGGTCGGAAGAGAGGAGATATACCAACCCCAATC 1243
Db 329 LeuAlaLysSerSerLysLysGlyValSerSerGluPheGlu-----ProPhe 345
Qy 1244 AACTTGAGAGAGCGGAGCCGATCTTCTTAACAACCTTGGGAGTTATTGAGGTGAAG 1303
Db 346 AsnLeuArgSerGlnAsnProLysTyrSerAsnLysPheGlyLysLeuPheGluLeuThr 365
Qy 1304 CCAGCAAGAGAAACCCCGCTTCAGGACCTGACATGATCTCACCTGTGTAGAGATC 1363
Db 366 ProGluLysLysTyrProGlnLeuGlnAspLeuAspLeuPheValSerSerValGluLeu 385
Qy 1364 AAAGAAGAGCTTTGATGCTCCCACTCACTCAAAAGCCATGTTATCGTGTGCTC 1423
Db 386 AsnGluGlyGlyLeuMetLeuProHisTyrAsnSerArgAlaIleValIleLeuLeuVal 405
Qy 1424 AACAAAGGAACTGGAAACCTTGAACCTGCTGCTTAAGAAAGAGCAACACAGAGGGGA 1483
Db 406 AsnGluGlyLysGlyAsnLeuGluLeuValGlyLeuLysAsnGluGlnGlnArg 425
Qy 1484 CGCGGGAAGAAGAGGAGGAGCAAGACGAAGAGAGGAGGAAGTAACAGAGAGGTGCGT 1543
Db 426 GluArgGluAspGluGln-----GlnValGln 434
Qy 1544 AGGTACACAGCGAGGTTGAAGAGGCGATGTGTTCATCATGCCAGCAGCTATCCAGTA 1603
Db 435 ArgTyrGluAlaArgLeuSerProGlyAspValValIleIleProAlaGlyHisProVal 454
Qy 1604 GCCATCAACGCTTCTCGAATCCATCTGCTGGCTTCGGTATCAACGCTGAACCAAC 1663
Db 455 AlaValSerAlaSerSerAsnLeuAsnLeuLeuGlyPheGlyIleAsnAlaGluAsnAsn 474
Qy 1664 CACAGAATCTTCTTCGAGGTGATAAGGACAATGTGATACAGACAGATAGAGCAAGCG 1723
Db 475 GlnArgAsnPheLeuThrGlySerAspAsnValIleSerGlnIleGluAsnProVal 494
Qy 1724 AAGATTATGATTCCTCGGTGGGTGAACAAAGTTGAGAGCTCATCAAAACCAAGAG 1783
Db 495 LysGluLeuThrPheProGlySerAlaGlnGluValAsnArgLeuLeuLysAsnGlnGlu 514
Qy 1784 GAATCTCACTTGTGAGTGTCTGCTCAATCTCAATCTCAATCTCGCTGCTCTGAG 1843
Db 515 HisSerHisPheAlaAsnAlaGluPro----- 523
Qy 1844 AAAGAGTCTCTGAGAAAGAGGATCAAGAGAGGAGAAACCAAGAGGAGGTCCTC 1903
Db 524 -----GluGlnLysGlyGluGluSerGlnArgLysArgSerProIle 537
Qy 1904 CTTTCAATTTGAAGGCTTTTAAC 1927
Db 538 SerSerIleLeuGlyThrPheAsn 545

RESULT 15
Q9M3X6 PRELIMINARY; PRT; 613 AA.
ID Q9M3X6 AC Q9M3X6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Convicillin precursor.
GN Name=cvc;
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
OX NCB1_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Seed cotyledon;
RA Casey R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
```

```
DR EMBL; AJ276875; CAB82855.1; -.
DR HSSP; P25974; IIPJ.
DR GO; GO:0045735; F:nutrient reservoir activity; IEA.
DR InterPro; IPR006045; Cupin.
DR InterPro; IPR007113; Cupin region.
DR InterPro; IPR011051; RmlC_like_cupin.
DR Pfam; PF00190; Cupin; 2.
KW Signal.
FT SIGNAL 1 30 Potential.
FT CHAIN 30 613 convicillin.
SQ SEQUENCE 613 AA; 72063 MW; 49AED99F6135DD19 CRC64;

Alignment Scores:
Pred. No.: 1.14e-73 Length: 613
Score: 1251.00 Matches: 283
Percent Similarity: 57.61% Conservative: 107
Best Local Similarity: 41.80% Mismatches: 155
Query Match: 35.12% Indels: 132
DB: 2 Gaps: 17

US-10-728-051-1 (1-2032) x Q9M3X6 (1-613)
Qy 68 CCACTGATGCTGTGTAGGATCCTGTCTGGCTTCA---CTTTCTGCAACGATGCC 124
Db 10 ProLeuLeuLeuLeuGlyIleIlePheLeuAlaSerValValCysValThrTyrAla 29
Qy 125 AAGTCATCATCCTTACCAGAAAGAAACACAGAG---AACCCCTGCCCCAGAGGTGCTCCAG 181
Db 30 Asn-----TyrAspGluGlySerGluProArgValProAlaGlnArg---Glu 44
Qy 182 AGTTGTCAACAGGAACCGGATGACTTGAAGCAAAAGGCATGCGAGTCTCGCTCCACCAAG 241
Db 45 ArgGlyArgGlnGluGlyGluLysGluGluLysArgHisGlyGluTrpArgProSer--- 63
Qy 242 CTCGATGATGATCCTCGTGTCTGTATGATCTCTCGAGGACACACTGGCACCACCAACAA 301
Db 64 -----TyrGlu-----LysGluGlu 68
Qy 302 CGTTCCCTCCAGGGAGGGGACACCTGCGCCCAACCCGGAGACTACGATGATGACCGC 361
Db 69 AspGluGluGluGlnArgGluArgGlyArgGlnGluGlyGluLysGluLysArg 88
Qy 362 CGTCAACCCCGAAGAGAGAGGAGCCGATGGGGACCACT----- 403
Db 89 His-----GlyGluTrpArgProSerTyrGluLysGlnGluAsp 101
Qy 403 ----- 403
Db 102 GluGluGluLysGlnLysTyrArgTyrGlnArgGluLysGluAspGluGluLysGln 121
Qy 404 -----GGACCGAGGAG 415
Db 122 LysTyrGlnTyrGlnArgGluLysLysGluGlnLysGluValGlnProGlyArgGluArg 141
Qy 416 CGTGAAGAGAGAAAGACTGGAGACAAACCAAGAGAAGATTGGAGG----- 460
Db 142 TrpGluArgGluAspGluGlnValAspGluGluTrpArgGlySerGlnArgArg 161
Qy 461 -----CGACCAAGTCATCAGCAGCAGCAGCAAAATAAGG 493
Db 162 GluAspProGluGluArgAlaArgLeuArgHisArgGluGluArgThrLysArgAspArg 181
Qy 494 CCGAAGGAAGAGAGGAGAAACAAGAGTGGGGAACACACAGGTAGCCATGTGAGGGAAGAA 553
Db 182 ArgHisGlnArgGluGlyGluGlu-----GluArgSerSerGluSer 196
Qy 554 ACATCTCGGAACAACCTTTCTACTTCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 613
Db 197 GlnGluArgArgAsnProPheLeuPheLysSerAsnLysPheLeuThrLeuPheGluAsn 216
Qy 614 CAAAACGGTAGGATCCGGTCTCTGAGAGGTTTGACAAAGGTCAAGGAGGTTTCAGAT 673
Db 217 GluAsnGlyHisIleArgLeuGlnArgPheAspLysArgSerAspLeuPheGluAsn 236
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 8, 2005, 04:11:27 ; Search time 19.3806 Seconds
(without alignments)
5523.384 Million cell updates/sec

Title: US-10-728-051-2

Perfect score: 1265

Sequence: 1 gtcaccatactagtagcccc.....taaaagatcatgttttggtt 717

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 1027090

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2_1/USPTO.spool_p/US10728051/runat_07042005_125046_14072/app.query.fasta_1.4757
-DB=Issued Patents AA -QFMT=faetan -SUFFIX=rai -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=BITS -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=spect -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10728051 @CGN 1.122 @runat_07042005_125046_14072 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A COMB.pcp.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pcp.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pcp.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pcp.*
5: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pcp.*
6: /cgn2_6/ptodata/1/iaa/backfilees1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	842	66.6	157	4	US-09-106-872A-22
2	842	66.6	157	4	US-09-191-593-10
3	842	66.6	159	4	US-09-106-872A-2
4	827	65.4	157	4	US-09-191-593-21
5	243	19.2	158	2	US-08-618-911-4
6	243	19.2	158	3	US-08-938-675A-2
7	243	19.2	158	4	US-09-531-727-2
8	241.5	19.1	155	2	US-08-618-911-2
9	229	18.1	158	2	US-08-618-911-6
10	156	12.3	28	4	US-09-191-593-53
11	156	12.3	158	2	US-08-670-186-4
12	151	11.9	158	2	US-08-670-186-6

13	139	11.0	155	2	US-08-670-186-2	Sequence 2, Appli
14	139	11.0	184	1	US-08-453-924-3	Sequence 3, Appli
15	138	10.9	28	4	US-09-191-593-58	Sequence 58, Appl
16	136.5	10.8	164	4	US-09-627-536-13	Sequence 13, Appl
17	118	9.3	30	4	US-09-106-872A-16	Sequence 16, Appl
18	118	9.3	30	4	US-09-191-593-57	Sequence 57, Appl
19	106.5	8.4	174	4	US-09-645-593-7	Sequence 7, Appli
20	106.5	8.4	1088	4	US-09-233-857-13	Sequence 13, Appl
21	105.5	8.3	1099	3	US-09-442-100-2	Sequence 2, Appli
22	105.5	8.3	1099	4	US-08-939-106-2	Sequence 2, Appli
23	105.5	8.3	1099	4	US-09-442-103-2	Sequence 2, Appli
24	102	8.1	316	3	US-08-098-327B-31	Sequence 31, Appl
25	102	8.1	316	3	US-08-462-625-31	Sequence 31, Appl
26	100	7.9	17	4	US-09-191-593-55	Sequence 55, Appl
27	100	7.9	17	4	US-09-191-593-61	Sequence 61, Appl
28	100	7.9	17	4	US-09-191-593-62	Sequence 62, Appl
29	100	7.8	224	4	US-09-270-767-44901	Sequence 44901, A
30	99.5	7.9	148	4	US-09-083-852A-4	Sequence 4, Appli
31	99.5	7.9	148	4	US-09-083-852A-6	Sequence 6, Appli
32	99.5	7.9	148	4	US-09-489-674B-6	Sequence 6, Appli
33	97	7.7	185	4	US-09-252-991A-23490	Sequence 23490, A
34	95	7.5	32	1	US-08-158-704-1	Sequence 1, Appli
35	95	7.5	32	2	US-08-610-424B-1	Sequence 1, Appli
36	95	7.5	32	4	US-09-336-463-1	Sequence 1, Appli
37	95	7.5	32	4	US-09-191-593-1	Sequence 1, Appli
38	95	7.5	152	4	US-09-252-991A-19284	Sequence 19284, A
39	94	7.4	43	3	US-09-303-814-1	Sequence 1, Appli
40	94	7.4	500	4	US-09-265-630-13	Sequence 13, Appl
41	93.5	7.4	415	4	US-09-252-991A-18252	Sequence 18252, A
42	93	7.4	537	4	US-09-252-991A-27024	Sequence 27024, A
43	92.5	7.3	788	2	US-08-918-914-4	Sequence 4, Appli
44	91.5	7.2	1898	1	US-08-056-200-94	Sequence 94, Appl
45	91.5	7.2	1898	2	US-08-800-644-94	Sequence 94, Appl

ALIGNMENTS

RESULT 1
US-09-106-872A-22
; Sequence 22, Application US/09106872A
; Patent No. 6486311
; GENERAL INFORMATION:
; APPLICANT: Burks Jr., A. Wesley
; APPLICANT: Stanley, J. Steven
; APPLICANT: Cockrell, Gael
; APPLICANT: King, Nina E.
; APPLICANT: Sampson, Hugh A.
; APPLICANT: Helm, Ricki M.
; APPLICANT: Bannon, Gary A.
; TITLE OF INVENTION: Peanut Allergens and Methods
; FILE REFERENCE: HS 103 CIP
; CURRENT APPLICATION NUMBER: US/09/106,872A
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: PCT/US96/15222
; PRIOR FILING DATE: 1996-09-23
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Arachis hypogaea
US-09-106-872A-22

Alignment Scores:
Pred. No.: 8.73e-86
Score: 842.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 66.56%
DB: 4
Length: 157
Matches: 157
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

US-10-728-051-2 (1-717) x US-09-106-872A-22 (1-157)

QY 2 CTCACCATCTAGTACGCTCGCCCTTTTCCTCCTCGCTGCCACGCATCTCGAGGCGAG 61
Db 1 LeuThrIleLeuValAlaLeuAlaLeuPheLeuLeuAlaAlaHisAlaSerAlaArgGln 20
QY 62 CAGTGGGACTCCAAAGGACACAGAGATCCAGAGCCAGCTCGAGAGGCGGACCTGAGG 121
Db 21 GlnTrpGluGlnGlnGlyAspArgArGysGlnSerGlnLeuGluArgAlaAsnLeuArg 40
QY 122 CCCTGGGAGCAACATCTCATGCAGAGATCCAAACGTGACGAGGATTCATATCAACGGGAC 181
Db 41 ProCysGluGlnHisLeuMetGlnIleGlnArgAspGluAspSerTyrgluArgAsp 60
QY 182 CCCTACAGCCCTAGTCAGATCCGTCACAGCCCTAGTCATATATCGAGAGCGCTGGA 241
Db 61 ProTyrsSerProSerGlnAspProTyrsSerProSerProTyrsAspArgGlyAlaGly 80
QY 242 TCCTCTCAGCACCACAGAGGTTGTCGATGAGCTGACGAGTTCGAGTTCGAGAACCAAGG 301
Db 81 SerSerGlnHisGlnGluArgCysCysAsnGluLeuAsnGluPheGluAsnGlnArg 100
QY 302 TGATGTGGGAGGATTCGCAACAGATCATCGAGAGCCAGAGCGATAGTTCAGGGGAGG 361
Db 101 CysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArgLeuGlnArg 120
QY 362 CAACAGGAGCAACAGTTCAGAGGGAGCTCAGGAATTCCTCAACAGTGGCGGCTTAGG 421
Db 121 GlnGlnGlnGlnPheIleArgGluLeuArgAsnLeuProGlnGlnCysGlyLeuArg 140
QY 422 GCACACAGGTTGGCAGTTCGAGGTCGAAAGTGGCGGCGAGACAGATAC 472
Db 141 AlaProGlnArgCysAspLeuAspValGluSerGlyArgAspArgTyr 157

RESULT 2
US-09-191-593-10
; Sequence 10, Application US/09191593
; Patent No. 6835824
; GENERAL INFORMATION:
; APPLICANT: BURKS, A Wesley, HELM, Ricki M,
; APPLICANT: COCKRELL, Gael, STANLEY, J Steven,
; APPLICANT: BANNON, Gary A
; TITLE OF INVENTION: PEANUT ALLERGENS AND
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Head, Johnson & Kachigian
; STREET: 112 W. Center St., Suite 230
; CITY: Fayetteville
; STATE: Arkansas AR
; COUNTRY: United States of America
; ZIP: 72701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS 6.2
; SOFTWARE: Wordperfect 6.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09191,593
; FILING DATE: 13 NOVEMBER 1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/717,933
; FILING DATE: 23 SEPTEMBER 1996
; APPLICATION NUMBER: US 07/998,377
; FILING DATE: 30 DECEMBER 1992
; APPLICATION NUMBER: US 08/158,704
; FILING DATE: 29 NOVEMBER 1993
; APPLICATION NUMBER: US 60/009,455
; FILING DATE: 29 DECEMBER 1995
; APPLICATION NUMBER: US 08/610,424
; FILING DATE: 04 MARCH 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ALEXANDER, DANIEL R

; REGISTRATION NUMBER: 32,604
; REFERENCE/DOCKET NUMBER: ARK00895601B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (501) 582-9111
; TELEFAX: (501) 521-4931
; TELEX: No. 6835824 applicable
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 157 amino acids
; TYPE: amino acid
; STRANDEDNESS: not applicable
; TOPOLOGY: Unknown
; MOLECULE TYPE: cDNA
; DESCRIPTION: identified as Ara h II cDNA clone
; DESCRIPTION: derived amino acid sequence
; HYPOTHETICAL: NO
; ANTI-SENSE: No
; FRAGMENT TYPE: No. 6835824 applicable
; ORIGINAL SOURCE:
; ORGANISM: Arachis hypogaea
; STRAIN: Florunner
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE: seed
; HAPLOTYPE: No. 6835824 applicable
; TISSUE TYPE: seed cDNA
; CELL TYPE: No. 6835824 applicable
; CELL LINE: No. 6835824 applicable
; ORGANELLE: No. 6835824 applicable
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: No. 6835824 applicable
; MAP POSITION: No. 6835824 applicable
; UNITS: No. 6835824 applicable
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD: By agreement with
; IDENTIFICATION METHOD: protein information and established
; IDENTIFICATION METHOD: consensus sequence
; OTHER INFORMATION: Seed storage protein and
; OTHER INFORMATION: allergen
; PUBLICATION INFORMATION:
; AUTHORS:
; TITLE:
; JOURNAL:
; VOLUME:
; ISSUE:
; PAGES:
; DATE:
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO:
; US-09-191-593-10
Alignment Scores:
Pred. No.: 8.73e-86 Length: 157
Score: 842.00 Matches: 157
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 66.56% Indels: 0
DB: 4 Gaps: 0
US-728-051-2 (1-717) x US-09-191-593-10 (1-157)
QY 2 CTCACCATCTAGTACGCTCGCCCTTTTCCTCCTCGCTGCCACGCATCTCGAGGCGAG 61
Db 1 LeuThrIleLeuValAlaLeuPheLeuLeuAlaAlaHisAlaSerAlaArgGln 20
QY 62 CAGTGGGAACTCCAAGGACACAGAGATCGCAGAGCGAGCTCGAGAGGCGGACCTGAGG 121

```
Db 21 GlnTrpGluLeuGlnGlyAspArgCysGlnSerGlnLeuGluArgAlaAsnLeuArg 40
QY 122 CCTGGAGCAACATCTCAGCAGAGATCCACGTCAGCAGGATTCATATGACGGGAC 181
Db 41 ProCysGluGlnHisLeuMetGlnLysIleGlnArgAspGluAspSerTyrGluArgAsp 60
QY 182 CCCTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCATATCATCGAGAGCGCTGGA 241
Db 61 ProTyrSerProSerGlnAspProTyrSerProSerProTyrAspArgArgGlyAlaGly 80
QY 242 TCCTCTCAGCACCAAGAGAGGTGTGCAATGAGCTGAACGAGTTTGAGAACACCAAGG 301
Db 81 SerSerGlnHisGlnGluArgCysCysAsnGluLeuAsnGluPheGluAsnAsnGlnArg 100
QY 302 TGATGTGCGAGGATTCGAACAGATCATCGAGAACAGAGCGATAGTTGCGGGGAGG 361
Db 101 CysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArgLeuGlnGlyArg 120
QY 362 CAACAGGAGCAACAGTTCAAGAGGGAGCTCAGGAACCTTGCTCAACAGTCGCGCTTAGG 421
Db 121 GlnGlnGluGlnGlnPheLysArgGluLeuArgAsnLeuProGlnGlnCysGlyLeuArg 140
QY 422 GCACCAAGGCTTGGACGTGCGACGTGCGAAAGTGGCGGCGAGACAGATAC 472
Db 141 AlaProGlnArgCysAspLeuAspValGluSerGlyArgAspArgTyr 157

RESULT 3
US-09-106-872A-2
; Sequence 2, Application US/09106872A
; Patent No. 6486311
; GENERAL INFORMATION:
; APPLICANT: Burks Jr., A. Wesley
; APPLICANT: Stanley, J. Steven
; APPLICANT: Cockrell, Gael
; APPLICANT: King, Nina E.
; APPLICANT: Sampson, Hugh A.
; APPLICANT: Helm, Ricki M.
; APPLICANT: Bannon, Gary A.
; TITLE OF INVENTION: Peanut Allergens and Methods
; FILE REFERENCE: HS 103 CIP
; CURRENT APPLICATION NUMBER: US/09/106,872A
; CURRENT FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: PCT/US96/15222
; PRIOR FILING DATE: 1996-09-23
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Arachis hypogaea
; FEATURE:
; OTHER INFORMATION: Amino Acids 17-26 are Ara H 2 Binding Epitopes,
; OTHER INFORMATION: Peptide 1
; OTHER INFORMATION: Amino Acids 23-32 are Ara H 2 binding epitopes,
; OTHER INFORMATION: peptide 2
; OTHER INFORMATION: Amino Acids 29-38 are Ara H 2 binding epitopes,
; OTHER INFORMATION: peptide 3
; OTHER INFORMATION: Amino Acids 41-50 are Ara H 2 binding epitopes,
; OTHER INFORMATION: peptide 4
; OTHER INFORMATION: Amino Acids 51-60 are Ara H 2 binding epitopes,
; OTHER INFORMATION: peptide 5
; OTHER INFORMATION: Amino Acids 59-68 are Ara H 2 binding epitopes,
; OTHER INFORMATION: peptide 6
; OTHER INFORMATION: Amino Acids 67-76 are Ara H 2 binding epitopes,
; OTHER INFORMATION: peptide 7
; OTHER INFORMATION: Amino Acids 117-126 are Ara H 2 binding epitopes,
; OTHER INFORMATION: peptide 8
; OTHER INFORMATION: Amino Acids 129-138 are Ara H 2 binding epitopes,
; OTHER INFORMATION: peptide 9
; OTHER INFORMATION: Amino Acids 145-154 are Ara H 2 binding epitopes,
; OTHER INFORMATION: peptide 10
; OTHER INFORMATION: Amino Acids 13-22 are Ara H 2 binding epitopes,
; OTHER INFORMATION: peptide 1, Table 12
```

```
; OTHER INFORMATION: Amino Acids 27-36 are Ara H 2 binding epitopes,
; OTHER INFORMATION: peptide 2, Table 12
; OTHER INFORMATION: Amino Acids 59-68 are Ara H 2 binding epitopes,
; OTHER INFORMATION: peptide 3, Table 12
; OTHER INFORMATION: Amino Acids 81-90 are Ara H 2 binding epitopes,
; OTHER INFORMATION: peptide 4, Table 12
; OTHER INFORMATION: Amino Acids 91-100 are Ara H 2 binding epitopes,
; OTHER INFORMATION: peptide 5, Table 12
; OTHER INFORMATION: Amino Acids 145-154 are Ara H 2 binding epitopes,
; OTHER INFORMATION: peptide 6, Table 12
US-09-106-872A-2
```

Alignment Scores:

Pred. No.:	8,77e-86	Length:	159
Score:	842.00	Matches:	157
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	66.56%	Indels:	0
DB:	4	Gaps:	0

US-10-728-051-2 (1-717) x US-09-106-872A-2 (1-159)

```
QY 2 CTCACCATACTAGTAGCCCTCGCCCTTTTCCTCTCGTCCGCCACGCATCTCGAGGCGAG 61
Db 3 LeuThrIleLeuValAlaLeuAlaLeuPheLeuLeuAlaAlaHisAlaSerAlaArgGln 22
QY 62 CAGTGGAACTCCAAGGAGACAGATGCCAGAGCCAGCTCGAGAGGCGGCAACCTGAGG 121
Db 23 GlnTrpGluLeuGlnGlyAspArgCysGlnSerGlnLeuGluArgAlaAsnLeuArg 42
QY 122 CCCTGGAGCAACATCTCATGCAAGATCCACGTCGAGAGGATTCATATGACCGGAC 181
Db 43 ProCysGluGlnHisLeuMetGlnLysIleGlnArgAspGluAspSerTyrGluArgAsp 62
QY 182 CCGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCATATCATCGAGAGCGCTGGA 241
Db 63 ProTyrSerProSerGlnAspProTyrSerProSerProTyrAspArgArgGlyAlaGly 82
QY 242 TCCTCTCAGCACCAAGAGAGGTGTGCAATGAGCTGAACGAGTTTGAGAACACCAAGG 301
Db 83 SerSerGlnHisGlnGluArgCysCysAsnGluLeuAsnGluPheGluAsnAsnGlnArg 102
QY 302 TGATGTGCGAGGATTCGAACAGATCATCGAGAACAGAGCGATAGTTGCGAGGAGG 361
Db 103 CysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArgLeuGlnGlyArg 122
QY 362 CAACAGGAGCAACAGTTCAAGAGGGAGCTCAGGAACCTTGCTCAACAGTCGCGCTTAGG 421
Db 123 GlnGlnGluGlnGlnPheLysArgGluLeuArgAsnLeuProGlnGlnCysGlyLeuArg 142
QY 422 GCACCAAGGCTTGGACGTGCGACGTGCGAAAGTGGCGGCGAGACAGATAC 472
Db 143 AlaProGlnArgCysAspLeuAspValGluSerGlyArgAspArgTyr 159
```

RESULT 4

```
US-09-191-593-21
; Sequence 21, Application US/09191593
; Patent No. 6835824
; GENERAL INFORMATION:
; APPLICANT: BURKS, A Wesley, HELM, Ricki M,
; APPLICANT: COCKRELL, Gael, STANLEY, J Steven,
; APPLICANT: BANNON, Gary A
; TITLE OF INVENTION: PEANUT ALLERGENS AND
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Head, Johnson & Kachigian
; STREET: 112 W. Center St., Suite 230
; CITY: Fayetteville
; STATE: Arkansas AR
; COUNTRY: United States of America
; ZIP: 72701
; COMPUTER READABLE FORM:
```

MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS-DOS 6.2
SOFTWARE: Wordperfect 6.0C
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/191,593
FILING DATE: 13 NOVEMBER 1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/717,933
FILING DATE: 23 SEPTEMBER 1996
APPLICATION NUMBER: US 07/998,377
FILING DATE: 30 DECEMBER 1992
APPLICATION NUMBER: US 08/158,704
FILING DATE: 29 NOVEMBER 1993
APPLICATION NUMBER: US 60/009,455
FILING DATE: 29 DECEMBER 1995
APPLICATION NUMBER: US 08/610,424
FILING DATE: 04 MARCH 1996
ATTORNEY/AGENT INFORMATION:
NAME: ALEXANDER, DANIEL R
REGISTRATION NUMBER: 32,604
REFERENCE/DOCKET NUMBER: ARK00895601B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (501) 582-9111
TELEFAX: (501) 521-4931
TELEX: No. 6835824 applicable
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 157 amino acids
TYPE: amino acid
STRANDEDNESS: No. 6835824 applicable
TOPOLOGY: unknown
MOLECULE TYPE: glycoprotein
DESCRIPTION: identified as Ara h 2 P38 deduced
DESCRIPTION: sequence from nucleotide sequence reading frame.
HYPOTHETICAL: No
ANTI-SENSE: No. 6835824 applicable
FRAGMENT TYPE: No. 6835824 applicable
ORIGINAL SOURCE:
ORGANISM: Arachis hypogaea
STRAIN: Florunner
INDIVIDUAL ISOLATE: Ara h II
DEVELOPMENTAL STAGE:
HAPLOTYPE: No. 6835824 applicable
TISSUE TYPE:
CELL TYPE: No. 6835824 applicable
CELL LINE: No. 6835824 applicable
ORGANELLE: No. 6835824 applicable
IMMEDIATE SOURCE:
LIBRARY:
CLONE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT: No. 6835824 applicable
MAP POSITION: No. 6835824 applicable
UNITS: No. 6835824 applicable
FEATURE:
NAME/KEY: final check
LOCATION: 4..480
IDENTIFICATION METHOD: By agreement with
IDENTIFICATION METHOD: protein information and established
IDENTIFICATION METHOD: consensus sequence
OTHER INFORMATION: Seed storage protein and
OTHER INFORMATION: allergen
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:

DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-09-191-593-21
Alignment Scores:
Pred. No.: 4.19e-84 Length: 157
Score: 827.00 Matches: 155
Percent Similarity: 98.73% Conservative: 0
Best Local Similarity: 98.73% Mismatches: 2
Query Match: 65.38% Indels: 0
DB: 4 Gaps: 0
US-10-728-051-2 (1-717) x US-09-191-593-21 (1-157)
QY 2 CTCACCATAGTAGCCCTCGCCCTTTTCCTCCGTCGCCACGCATCTCGGAGGAG 61
Db 1 LeuThrIleLeuValAlaLeuPheLeuLeuAlaHisAlaSerAlaArgGln 20
QY 62 CAGTGGGAACCTCAAAGGAGACAGAGATGCCAGAGCCAGCTCGAGAGGGCGAACCTGAGG 121
Db 21 GlnTrpGluLeuGlnGlyAspArgCysGlnSerGlnLeuGluArgAlaAsnLeuArg 40
QY 122 CCCTCGAGCAACATCTCATGCAGAAAGATCCAACTGAGAGGATTCATATGAACGGGAC 181
Db 41 ProCysGluGlnHisLeuMetGlnLysIleGlnArgAspGluAspSerTyrGluArgAsp 60
QY 182 CGGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCATATGATCGGAGAGCGCTGGA 241
Db 61 ProSerTyrProSerGlnAspProTyrSerProSerProTyrAspArgGlyAlaGly 80
QY 242 TCCTCTCAGCACCAAGAGAGAGGTGTTGCAATGAGCTGAAAGAGTTTGAGAACCAACAAAGG 301
Db 81 SerSerGlnHisGlnGluArgCysAsnGluLeuAsnGluPheGluAsnAsnGlnArg 100
QY 302 TGCATGTGGAGGCAATTCACAGATCATATGAGAACCAAGAGAGGATAGTTGAGGGGAGG 361
Db 101 CysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArgLeuGlnGlyArg 120
QY 362 CAACAGGAGCAACAGTTCAAGAGGGAGGCTCAGGAACCTTGCCTCAACAGTGGCGGCTTAGG 421
Db 121 GlnGlnGluGlnGlnPheLysArgGluLeuArgAsnLeuProGlnGlnCysGlyLeuArg 140
QY 422 GCACACAGCGTTGGGACTTGGACGTGCAAAAGTGGCGGAGAGACAGATAC 472
Db 141 AlaProGlnArgCysAspLeuAspValGluSerGlyGlyArgAspArgTyr 157
RESULT 5
US-08-618-911-4
Sequence 4, Application US/08618911
Patent No. 5850016
GENERAL INFORMATION:
APPLICANT: Jung, Rudolf
APPLICANT: Hastings, Craig
APPLICANT: Coughlan, Sean
APPLICANT: Hu, David
TITLE OF INVENTION: ALTERATION OF AMINO ACID COMPOSITIONS IN
TITLE OF INVENTION: SEEDS
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.
STREET: 700 Capital Square, 400 Locust Street
CITY: Des Moines
STATE: Iowa
COUNTRY: USA
ZIP: 50309
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:


```
; APPLICATION NUMBER: US/08/618,911
; FILING DATE: Concurrently herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Simon, Soma
; REGISTRATION NUMBER: 37,444
; REFERENCE/DOCKET NUMBER: 365-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (515) 248-4800
; TELEFAX: (515) 248-4844
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 158 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-618-911-4

Alignment Scores:
Pred. No.: 1,26e-18 Length: 158
Score: 243.00 Matches: 56
Percent Similarity: 56.49% Conservative: 31
Best Local Similarity: 36.36% Mismatches: 55
Query Match: 19.21% Indels: 12
DB: 2 Gaps: 6

US-10-728-051-2 (1-717) x US-08-618-911-4 (1-158)
QY 5 ACCATACTAGTACCCCTCGCCCTTTCTCTCGTCGCCACGCATCTCGAGGCGACG 64
Db 5 ThrIleLeuLeuSerLeuLeuPheCysIle---AlaHisThrCysSerAlaSerLys 23
QY 65 TGGGAACCTCAAGGAGACAGAGATGCCAGAGCCAGCTCGAGGGCGGAACTGAGGCC 124
Db 24 TrpGlnHisGlnGlnAep---SerCysArgLysGlnLeuGlnGlyValAsnLeuThrPro 42
QY 125 TCGGACCAACATCTCATGAGAGATCCAA-----CGTGACGAGGATTCATATGAACGG 178
Db 43 CysGluLysHisIleMetGluLysIleGlnGlyArgGlyAspAspAspAsp 62
QY 179 GACCCGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCATATGATCGAGAGCGCT 238
Db 63 AspAspAsnHisIleLeuArgThrMetArgGlyArgIleAsnTyrlleArgAsnGlu 82
QY 239 GGATCTCTCAG-----CACCAGAGAGGTGTTCATATGAGCTGAAC 280
Db 83 GlyLysAspGluAspGluGluGluGlyHisMetGlnLysCysCysThrGluMetSer 102
QY 281 GAGTTTGAGAACCAACCAAGGTGCTGCGAGGCATTCGAGAGGATTCGAGAGAACCG 340
Db 103 GluLeu---ArgSerProLysCysGlnCysLysAlaLeuGlnLysIleMetGluAsnGln 121
QY 341 AGCGATAGTTGCGAGGGGCGGCAACAGGACAGTTCAGAGGGAGCTCAGGAACCTTG 400
Db 122 SerGluGluLeuGlu---GluLysGlnLysLysMetGluLysGluLeuIleAsnLeu 140
QY 401 CCTCAACAGTGGCGCTTAGGGCACCAACAGCGTTGCGACTTG 442
Db 141 AlaThrMetCysArgPheGlyProMetIleGlnCysAspLeu 154

RESULT 6
US-08-938-675A-2
; Sequence 2, Application US/08938675A
; Patent No. 6107287
; GENERAL INFORMATION:
; APPLICANT: de Lumen, Benito O.
; APPLICANT: Galvez, Alfredo F.
; TITLE OF INVENTION: Lunasin Peptides
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA

; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,675A
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: OSWAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B98-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 158 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-938-675A-2

Alignment Scores:
Pred. No.: 1,26e-18 Length: 158
Score: 243.00 Matches: 56
Percent Similarity: 56.49% Conservative: 31
Best Local Similarity: 36.36% Mismatches: 55
Query Match: 19.21% Indels: 12
DB: 3 Gaps: 6

US-10-728-051-2 (1-717) x US-08-938-675A-2 (1-158)
QY 5 ACCATACTAGTACCCCTCGCCCTTTCTCTCGTCGCCACGCATCTCGAGGCGACG 64
Db 5 ThrIleLeuLeuSerLeuLeuPheCysIle---AlaHisThrCysSerAlaSerLys 23
QY 65 TGGGAACCTCAAGGAGACAGAGATGCCAGAGCCAGCTCGAGGGCGGAACTGAGGCC 124
Db 24 TrpGlnHisGlnGlnAep---SerCysArgLysGlnLeuGlnGlyValAsnLeuThrPro 42
QY 125 TCGGACCAACATCTCATGAGAGATCCAA-----CGTGACGAGGATTCATATGAACGG 178
Db 43 CysGluLysHisIleMetGluLysIleGlnGlyArgGlyAspAspAspAsp 62
QY 179 GACCCGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCATATGATCGAGAGCGCT 238
Db 63 AspAspAsnHisIleLeuArgThrMetArgGlyArgIleAsnTyrlleArgAsnGlu 82
QY 239 GGATCTCTCAG-----CACCAGAGAGGTGTTCATATGAGCTGAAC 280
Db 83 GlyLysAspGluAspGluGluGluGlyHisMetGlnLysCysCysThrGluMetSer 102
QY 281 GAGTTTGAGAACCAACCAAGGTGCTGCGAGGCATTCGAGAGGATTCGAGAGAACCG 340
Db 103 GluLeu---ArgSerProLysCysGlnCysLysAlaLeuGlnLysIleMetGluAsnGln 121
QY 341 AGCGATAGTTGCGAGGGGCGGCAACAGGACAGTTCAGAGGGAGCTCAGGAACCTTG 400
Db 122 SerGluGluLeuGlu---GluLysGlnLysLysMetGluLysGluLeuIleAsnLeu 140
QY 401 CCTCAACAGTGGCGCTTAGGGCACCAACAGCGTTGCGACTTG 442
Db 141 AlaThrMetCysArgPheGlyProMetIleGlnCysAspLeu 154

RESULT 7
US-09-531-727-2
; Sequence 2, Application US/09531727
; Patent No. 6544956
```

GENERAL INFORMATION:
APPLICANT: de Lumen, Benito O.
Galvez, Alfredo F.
TITLE OF INVENTION: Lunasin Peptides
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 75 DENISE DRIVE
CITY: HILLSBOROUGH
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94010
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/531,727
FILING DATE: 21-Mar-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/938,675
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: B98-003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 158 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-531-727-2

Alignment Scores:
Pred. No.: 1,26e-18 Length: 158
Score: 243.00 Matches: 56
Percent Similarity: 56.49% Conservative: 31
Best Local Similarity: 36.36% Mismatches: 55
Query Match: 19.21% Indels: 12
DB: 4 Gaps: 6

US-10-728-051-2 (1-717) x US-09-531-727-2 (1-158)

QY 5 ACCATAGTAGCCCTCGCCCTTTCTCTCGTCCAGCATCTGCGAGGCAGCAG 64
Db 5 ThriLeuLeuLeuSerLeuPheCysIle---AlaHisThrCysSerAlaSerLys 23
QY 65 TGGGAACCTCAAGAGAGACAGAGATCCAGAGCCAGCTCGAGAGGCGCAACTGAGGCC 124
Db 24 TrpGlnHisGlnAsp---SerCysArgLysGlnLeuGlnGlyValAsnLeuThrPro 42
QY 125 TGGAGCAACATCTATCGACAGATCCAA-----CGTGACGAGGATTCATATGAACGG 178
Db 43 CysGluYshHisIleMetGluLysIleGlnGlyArgGlyAspAspAspAspAsp 62
QY 179 GACCCGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCATATCATCGAGAGCGCT 238
Db 63 AspAspAsnHisIleLeuArgThrMetArgGlyArgIleAsnTyrlleArgArgAsnGlu 82
QY 239 GGATCCTCTCAG-----CACCAAGAGAGGTGTTCATATGAGCTGAAC 280
Db 83 GlyLysAspGluAspGluGluGluGluGluGlyHisMetGlnLysCysCysThrGluMetSer 102
QY 281 GAGTTTGAGAACACCAAGGTGCATGTGCGAGCATTCGACAGATTCGAGAACCG 340
Db 281 GAGTTTGAGAACACCAAGGTGCATGTGCGAGCATTCGACAGATTCGAGAACCG 340

Db 103 GluLeu---ArgSerProLysCysGlnCysLysAlaLeuGlnLysIleMetGluAsnGln 121
QY 341 AGCATAGGTTCAGGAGGAGCAACAGGAGCAACAGATTCAAGAGGAGCTCAGGAACCTTG 400
Db 122 SerGluGluLeuGlu---GluLysGlnLysLysMetGluLysGluLeuIleAsnLeu 140
QY 401 CCTCAACAGTGGCGCTTAGGACACACAGGCTTCGAGCTTG 442
Db 141 AlaThrMetCysArgPheGlyProMetIleGlnCysAspLeu 154
RESULT 8
US-08-618-911-2
Sequence 2, Application US/08618911
GENERAL INFORMATION:
APPLICANT: Jung, Rudolf
APPLICANT: Hastings, Craig
APPLICANT: Coughlan, Sean
APPLICANT: Hu, David
TITLE OF INVENTION: ALTERATION OF AMINO ACID COMPOSITIONS IN SEEDS
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.
STREET: 700 Capital Square, 400 Locust Street
CITY: Des Moines
STATE: Iowa
COUNTRY: USA
ZIP: 50309
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/618,911
FILING DATE: Concurrently herewith
ATTORNEY/AGENT INFORMATION:
NAME: Simon, Soma
REGISTRATION NUMBER: 37,444
REFERENCE/DOCKET NUMBER: 365-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (515) 248-4800
TELEFAX: (515) 248-4844
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 155 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-618-911-2

Alignment Scores:
Pred. No.: 1,85e-18 Length: 155
Score: 241.50 Matches: 58
Percent Similarity: 57.14% Conservative: 30
Best Local Similarity: 37.66% Mismatches: 53
Query Match: 19.09% Indels: 13
DB: 2 Gaps: 7

US-10-728-051-2 (1-717) x US-08-618-911-2 (1-155)

QY 2 CTCACCATAGTAGCCCTCGCCCTTTCTCTCGTCCAGCATCTGCGAGGCAG 61
Db 4 LeuThrIleLeuLeuAlaLeuPheIle-----AlaHisThrCysCysAlaSer 21
QY 62 CAGTGGAACTCAAGGACAGAGATCCAGAGCCAGCTCGAGAGGCGCAACTGAGG 121
Db 22 LysTrpGlnGlnHisGlnGlnGluSerCysArgGluGlnLeuLysGlyIleAsnLeuAsn 41
QY 122 CCTCGGACCAACATCTCATGACAGATCCAA-----CGTGACGAGGATTCATAT 172
Db 42 ProCysGlu---HisIleMetGluLysIleGlnAlaGlyArgArgGlyGluAspGlySer 60

```
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-670-186-2

Alignment Scores:
Pred. No.: 5,73e-07 Length: 155
Score: 139.00 Matches: 45
Percent Similarity: 45.51% Conservative: 26
Best Local Similarity: 28.85% Mismatches: 47
Query Match: 10.99% Indels: 38
DB: 2 Gaps: 9

US-10-728-051-2 (1-717) x US-08-670-186-2 (1-155)
QY 8 ATACTAGTACGCTCGCCCTTTCTCTCGCTGCCACGATCTCGGAGG----- 58
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7 LeuPheAlaThrLeuAlaLeuPheValLeuLeuAlaAsnAlaSerIleGlnThrVal 26
QY 59 -----CAGAGTGGGAATCCCAAGGAGACAGAAATGC 91
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27 ValGluValAspGluGluGluAspAsnGlnLeuTrp-----ArgCys 40
QY 92 CAGAGCCAG--CTCGAGAGGGCGAACCCTGAGGCGCTCGGAGCAACATCTCATGCAGAG 148
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
41 GlnArgGlnPheLeuGlnHisGlnArgLeuArgAlaCysGlnArgPheIleHisArg 60
QY 149 ATCCAA-----CGTGACGAGGATTTCATATGAACGGGACCGGTACAGCCCTAGTCAGGAT 202
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 AlaGlnPheGlyGlnProAspGluLeuGluAspGluValGluAspAsnAspAsp 80
QY 203 CCGTACAGCCCTAGTCCATATGATCGGAGGCGCTGGATCTCTCAGCAACAGAGG 262
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
81 GluAsnGlnPro-----ArgArgProAla-----LeuArgGln 91
QY 263 TGTTCGAATGAGTGAACGAGTTTGAGAACCAACAAAGGTGCATGTCC----- 310
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
92 CysCysAsnGlnLeuArgGlnValAsp-----ArgProCysValCysProValLeuArg 109
QY 311 GAGCGATTCACAGATCATCGAGAACCAAGGATAGGTTCGAGGGGAGGCAACAGGAG 370
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
110 GlnAlaAlaGlnValLeuGlnArgGlnIle-----IleGlnGlyProGlnGlnLeu 127
QY 371 CAACAGTTCAAGGGAGTTCAGGAATTCCTCCCTCAACAGTGGCGCTT 418
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
128 ArgArgLeuPheAspAlaAlaArgAsnLeuProAsnIleCysAsnIle 143

RESULT 14
US-08-453-924-3
; Sequence 3, Application US/08453924
; Patent No. 5608152
; GENERAL INFORMATION:
; APPLICANT: Kridl, Jean C.
; APPLICANT: Knauf, Vic C.
; TITLE OF INVENTION: Seed-Specific Transcriptional Regulation
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weil, Gotshal & Manges
; STREET: 2882 Sand Hill Road, Ste. 280
; CITY: Menlo Park
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,924
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/742,834
```

RESULT 15

US-09-191-593-58
; Sequence 58, Application US/09191593
; Patent No. 6835824
; GENERAL INFORMATION:
; APPLICANT: BURKS, A Wesley, HELM, Ricki M,
; APPLICANT: COCKRELL, Gael, STANLEY, J Steven,
; APPLICANT: BANNON, Gary A
; TITLE OF INVENTION: PEANUT ALLERGENS AND
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Head, Johnson & Kachigian
; STREET: 112 W. Center St., Suite 230
; CITY: Fayetteville
; STATE: Arkansas AR
; COUNTRY: United States of America
; ZIP: 72701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.2
; SOFTWARE: Wordperfect 6.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/191,593
; FILING DATE: 13 NOVEMBER 1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/717,933
; FILING DATE: 23 SEPTEMBER 1996
; APPLICATION NUMBER: US 07/998,377
; FILING DATE: 30 DECEMBER 1992
; APPLICATION NUMBER: US 08/158,704
; FILING DATE: 29 NOVEMBER 1993
; APPLICATION NUMBER: US 60/009,455
; FILING DATE: 29 DECEMBER 1995
; APPLICATION NUMBER: US 08/610,424
; FILING DATE: 04 MARCH 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ALEXANDER, DANIEL R
; REGISTRATION NUMBER: 32,604
; REFERENCE/DOCKET NUMBER: ARK00895601B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (501) 582-9111
; TELEFAX: (501) 521-4931
; TELEX: No. 6835824 applicable
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: amino acid
; STRANDEDNESS: No. 6835824 applicable
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; DESCRIPTION: identified as 17.5 kD N-terminal
; DESCRIPTION: sequence of Ara h 2 p38 (gene sequence 20..47).
; HYPOTHETICAL: No
; ANTI-SENSE: No. 6835824 applicable
; FRAGMENT TYPE: amino terminus
; ORIGINAL SOURCE:
; ORGANISM: Arachis hypogaea
; STRAIN: Florunner
; INDIVIDUAL ISOLATE: P38
; DEVELOPMENTAL STAGE:
; HAPLOTYPE: No. 6835824 applicable
; TISSUE TYPE:
; CELL TYPE: No. 6835824 applicable
; CELL LINE: No. 6835824 applicable
; ORGANELLE: No. 6835824 applicable
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE: P38

; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: No. 6835824 applicable
; MAP POSITION: No. 6835824 applicable
; UNITS: No. 6835824 applicable
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD: By agreement with
; IDENTIFICATION METHOD: protein information and established
; IDENTIFICATION METHOD: consensus sequence
; OTHER INFORMATION: Seed storage protein and
; OTHER INFORMATION: allergen
; PUBLICATION INFORMATION:

; AUTHORS:
; TITLE:
; JOURNAL:
; VOLUME:
; ISSUE:
; PAGES:
; DATE:
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO:
; US-09-191-593-58

Alignment Scores:
Pred. No.: 3.95e-07 Length: 28
Score: 138.00 Matches: 26
Percent Similarity: 92.86% Conservative: 0
Best Local Similarity: 92.86% Mismatches: 2
Query Match: 10.91% Indels: 0
DB: 4 Gaps: 0

US-10-728-051-2 (1-717) x US-09-191-593-58 (1-28)

QY	59	CAGCAGTGGAACTCCACGAGACACAGATGCCAGAGCCAGCTCCAGAGGGCGACCTG	118
Db	1	GlnGlnTrpGluLeuGlnGlyAspArgArgGlnSerGlnLeuGluArgAlaAsnLeu	20
QY	119	AGGCCCTCGAGCAACATCTCATG	142
Db	21	ArgProCysGluGlnHisLysMet	28

Search completed: April 8, 2005, 05:34:11
Job time : 22.3806 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 8, 2005, 05:25:59 ; Search time 82.3048 Seconds
(without alignments)
5784.401 Million cell updates/sec

Title: US-10-728-051-2
Perfect score: 1265
Sequence: 1 gctcaccatactagtagcccc.....taaaagatcatgttttgtt 717

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1418010 seqs, 331997259 residues

Total number of hits satisfying chosen parameters: 2836020

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DBV=xlp
-Q=/cgn2_1/USPTO.spool.p/US10728051/runat_07042005_125048_14128/app.query.fasta_1.4757
-DB=Published Applications AA -QFMT=fastaa -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10728051.cgn 1 1 663 @runat_07042005_125048_14128
-NCPU=6 -ICPU=3 -NO MAP= -LARGOQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=120 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

Alignment Scores: 9.72e-80 Length: 157
Pred. No.: 842.00 Matches: 157
Score:

No.	Score	Match	Length	DB	ID	Description
1	842	66.6	157	14	US-10-228-806-4	Sequence 4, Appl
2	842	66.6	157	15	US-10-100-303A-63	Sequence 63, Appl
3	835	66.0	160	17	US-10-899-551-4	Sequence 4, Appl
4	819	64.7	156	15	US-10-245-871-53	Sequence 53, Appl
5	819	64.7	156	15	US-10-253-286-53	Sequence 53, Appl
6	771	60.9	166	15	US-10-100-303A-82	Sequence 82, Appl
7	760	60.1	167	17	US-10-899-551-56	Sequence 56, Appl
8	699	55.3	166	9	US-09-731-221-77	Sequence 77, Appl
9	463	36.6	83	15	US-10-100-303A-81	Sequence 81, Appl
10	243	19.2	158	15	US-10-302-633-2	Sequence 2, Appl
11	243	19.2	158	15	US-10-424-599-169225	Sequence 169225,
12	241.5	19.1	155	15	US-10-424-599-284372	Sequence 284372,
13	177.5	14.0	168	14	US-10-165-289A-2	Sequence 2, Appl
14	149.5	11.8	167	16	US-10-382-066-29	Sequence 29, Appl
15	144	11.4	169	14	US-10-165-289A-4	Sequence 4, Appl
16	139	11.0	184	11	US-09-782-130-3	Sequence 3, Appl
17	139	11.0	184	14	US-10-100-121-48	Sequence 48, Appl
18	129.5	10.2	150	15	US-10-442-174A-1	Sequence 1, Appl
19	127.5	10.1	151	15	US-10-424-599-167336	Sequence 167336,
20	127	10.0	22	14	US-10-228-806-67	Sequence 67, Appl
21	126	10.0	102	15	US-10-424-599-219338	Sequence 219338,
22	119	9.4	20	9	US-09-731-221-66	Sequence 66, Appl
23	118	9.2	261	15	US-10-291-265-715	Sequence 715, App
24	117	9.2	20	9	US-09-731-221-63	Sequence 63, Appl
25	116	9.2	20	9	US-09-731-221-60	Sequence 60, Appl
26	115	9.1	20	9	US-09-731-221-61	Sequence 61, Appl
27	113	8.9	20	9	US-09-731-221-65	Sequence 65, Appl
28	112.5	8.9	101	15	US-10-424-599-275916	Sequence 275916,
29	112	8.9	20	9	US-09-731-221-54	Sequence 54, Appl
30	112	8.9	20	9	US-09-731-221-59	Sequence 59, Appl
31	112	8.9	20	9	US-09-731-221-62	Sequence 62, Appl
32	112	8.9	20	9	US-09-731-221-64	Sequence 64, Appl
33	111	8.8	20	9	US-09-731-221-57	Sequence 57, Appl
34	111	8.8	20	9	US-09-731-221-74	Sequence 74, Appl
35	111	8.8	20	9	US-09-731-221-75	Sequence 75, Appl
36	111	8.8	20	9	US-09-731-221-76	Sequence 76, Appl
37	110	8.7	25	15	US-10-245-871-60	Sequence 60, Appl
38	110	8.7	25	15	US-10-253-286-60	Sequence 55, Appl
39	109	8.6	20	15	US-10-245-871-55	Sequence 55, Appl
40	109	8.6	20	15	US-10-253-286-55	Sequence 55, Appl
41	108	8.5	20	9	US-09-731-221-53	Sequence 53, Appl
42	108	8.5	20	9	US-09-731-221-56	Sequence 56, Appl
43	108	8.5	20	9	US-09-731-221-67	Sequence 67, Appl
44	108	8.5	20	9	US-09-731-221-68	Sequence 68, Appl
45	108	8.5	50	15	US-10-424-599-250084	Sequence 250084,

ALIGNMENTS

RESULT 1
US-10-228-806-4
; Sequence 4, Application US/10228806
; Publication No. US20030049237A1
; GENERAL INFORMATION:
; APPLICANT: Bannan, et al.
; TITLE OF INVENTION: Methods and Reagents for Decreasing Clinical Reactions
; TITLE OF INVENTION: to Allergy
; FILE REFERENCE: 2002834-0043
; CURRENT APPLICATION NUMBER: US/10/228, 806
; CURRENT FILING DATE: 2002-08-26
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Arachis hypogaea
US-10-228-806-4

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 66.56% Indels: 0
DB: 14 Gaps: 0

US-10-728-051-2 (1-717) x US-10-228-806-4 (1-157)

QY 2 CTCACCATAGTACGCTCGCCCTTTCTCTCTCGTCCCAACGCATCTGCGAGGAG 61
DB 1 LeuThrIleLeuValAlaLeuAlaLeuPheLeuLeuAlaAlaHisAlaSerAlaAraGln 20
QY 62 CAGTGGGAACCTCAAGAGACAGAAAGATGCCAGAGCAGCTCAGAGGGGGAACCTGAGG 121
DB 21 GlnTrpGluLeuGlnGlyAspArgCysGlnSerGlnLeuGluArgAlaAsnLeuArg 40
QY 122 CCCTGGAGCAACATCTCATGACAGATCCAAACGTGACGAGATTTCATATGAACGGGAC 181
DB 41 ProCysGluGlnHisLeuMetGlnIleGlnArgAspGluAspSerTyrGluArgAsp 60
QY 182 CCCTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCCATATGATCGGAGGCGCTGGA 241
DB 61 ProTyrSerProSerGlnAspProTyrSerProSerProTyrAspArgGlyAlaGly 80
QY 242 TCCTCTCAGCACCAAGAGGCTGTCGAATGAGCTGAACGAGTTTGAGAACACCAAGG 301
DB 81 SerSerGlnHisGlnLeuGlnPheLeuAsnGluLeuAsnGluPheGluAsnGlnArg 100
QY 302 TGCATGTGCGAGGCATTCACACAGATCATGAGAACCGATAGGTTGACAGGGGAGG 361
DB 101 CysMetCysGluAlaLeuGlnIleMetGluAsnGlnSerAspArgLeuGlnArg 120
QY 362 CAACAGGAGCAACAGTTCAAGAGGGAGCTCAGGAACCTTGCCTCAACAGTGCAGGCTTAGG 421
DB 121 GlnGlnGluGlnPheLysArgGluLeuArgAsnLeuProGlnCysGlyLeuArg 140
QY 422 GCACACAGGCTTGCAGCTTGCAGCTCGAAAGTGGGGGAGACAGATAC 472
DB 141 AlaProGlnArgCysAspLeuAspValGluSerGlyArgAspArgTyr 157

RESULT 2

US-10-100-303A-63

; Sequence 63, Application US/10100303A
; Publication No. US20030202980A1
; GENERAL INFORMATION:
; APPLICANT: Caplan, et al.
; TITLE OF INVENTION: Methods and Reagents for Decreasing Clinical Reaction
; TITLE OF INVENTION: to Allergy
; FILE REFERENCE: 2002834-0166
; CURRENT APPLICATION NUMBER: US/10/100,303A
; CURRENT FILING DATE: 2002-03-18
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Arachis hypogaea, Prot/Nucleo Ara h 2
US-10-100-303A-63

Alignment Scores:

Pred. No.: 9,72e-80 Length: 157
Score: 842.00 Matches: 157
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 66.56% Indels: 0
DB: 15 Gaps: 0

US-10-728-051-2 (1-717) x US-10-100-303A-63 (1-157)

QY 2 CTCACCATAGTACGCTCGCCCTTTCTCTCTCGTCCCAACGCATCTGCGAGGAG 61
DB 1 LeuThrIleLeuValAlaLeuAlaLeuPheLeuLeuAlaAlaHisAlaSerAlaAraGln 20
QY 62 CAGTGGGAACCTCAAGAGACAGAAAGATGCCAGAGCAGCTCAGAGGGGGAACCTGAGG 121

DB 21 GlnTrpGluLeuGlnGlyAspArgCysGlnSerGlnLeuGluArgAlaAsnLeuArg 40
QY 122 CCCTGGAGCAACATCTCATGACAGATCCAAACGTGACGAGATTTCATATGAACGGGAC 181
DB 41 ProCysGluGlnHisLeuMetGlnIleGlnArgAspGluAspSerTyrGluArgAsp 60
QY 182 CCCTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCCATATGATCGGAGGCGCTGGA 241
DB 61 ProTyrSerProSerGlnAspProTyrSerProSerProTyrAspArgGlyAlaGly 80
QY 242 TCCTCTCAGCACCAAGAGGCTGTCGAATGAGCTGAACGAGTTTGAGAACACCAAGG 301
DB 81 SerSerGlnHisGlnLeuGlnPheLeuAsnGluLeuAsnGluPheGluAsnGlnArg 100
QY 302 TGCATGTGCGAGGCATTCACACAGATCATGAGAACCGATAGGTTGACAGGGGAGG 361
DB 101 CysMetCysGluAlaLeuGlnIleMetGluAsnGlnSerAspArgLeuGlnArg 120
QY 362 CAACAGGAGCAACAGTTCAAGAGGGAGCTCAGGAACCTTGCCTCAACAGTGCAGGCTTAGG 421
DB 121 GlnGlnGluGlnPheLysArgGluLeuArgAsnLeuProGlnCysGlyLeuArg 140
QY 422 GCACACAGGCTTGCAGCTTGCAGCTCGAAAGTGGGGGAGACAGATAC 472
DB 141 AlaProGlnArgCysAspLeuAspValGluSerGlyArgAspArgTyr 157

RESULT 3

US-10-899-551-4

; Sequence 4, Application US/10899551
; Publication No. US20050063994A1
; GENERAL INFORMATION:
; APPLICANT: Burks, A. Wesley
; APPLICANT: Caplan, Michael J.
; APPLICANT: Sampson, Hugh A.
; APPLICANT: Howard, Sosin B.
; APPLICANT: Bottomly, Kim H.
; TITLE OF INVENTION: Methods and Reagents for Decreasing Clinical Reaction to Allergy
; FILE REFERENCE: 2002834-0233
; CURRENT APPLICATION NUMBER: US/10/899,551
; CURRENT FILING DATE: 2004-07-26
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 160
; TYPE: PRT
; ORGANISM: species Arachis hypogaea
US-10-899-551-4

Alignment Scores:

Pred. No.: 5.35e-79 Length: 160
Score: 835.00 Matches: 156
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 66.01% Indels: 0
DB: 17 Gaps: 0

US-10-728-051-2 (1-717) x US-10-899-551-4 (1-160)

QY 2 CTCACCATAGTACGCTCGCCCTTTCTCTCTCGTCCCAACGCATCTGCGAGGAG 61
DB 4 LeuThrIleLeuValAlaLeuAlaLeuPheLeuLeuAlaAlaHisAlaSerAlaAraGln 23
QY 62 CAGTGGGAACCTCAAGAGACAGAAAGATGCCAGAGCAGCTCAGAGGGGGAACCTGAGG 121
DB 24 GlnTrpGluLeuGlnGlyAspArgCysGlnSerGlnLeuGluArgAlaAsnLeuArg 43
QY 122 CCCTGGAGCAACATCTCATGACAGATCCAAACGTGACGAGATTTCATATGAACGGGAC 181
DB 44 ProCysGluGlnHisLeuMetGlnIleGlnArgAspGluAspSerTyrGluArgAsp 63
QY 182 CCCTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCCATATGATCGGAGGCGCTGGA 241

Db 64 ProTyrSerProSerGlnAspProTyrSerProSerProTyrAspArgArgGlyAlaGly 83
Qy 242 TCCTCTCAGACCAAGAGAGGTGTGCAATGAGCTGAACAGTTTGAGAACAAACCAAGG 301
Db 84 SerSerGlnHisGlnGluArgCysAsnGluLeuAsnGluPheGluAsnGlnArg 103
Qy 302 TGCATGTGCCAGGCACTTGCACAGATCATCGAGAACCCAGAGCGATAGTTTGCAGGGGAGG 361
Db 104 CysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArgLeuGlnGlyArg 123
Qy 362 CAACAGGAGCAACAGTTCAAGAGGAGCTCAGGAACCTTGCTCAACAGTCCGGCTTAGG 421
Db 124 GlnGlnGluGlnGlnPheLysArgGluLeuArgAsnLeuProGlnGlnCysGlyLeuArg 143
Qy 422 GCACCACAGCGTTGCCGACTTGGACCTGGAACCTCGAAGTGGCGGC 460
Db 144 AlaProGlnArgCysAspLeuAspValGluSerGlyGly 156
RESULT 4
US-10-245-871-53
; Sequence 53, Application US/10245871
; Publication No. US20030235594A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT
; APPLICANT: XU, MINZHEN
; TITLE OF INVENTION: II-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: REH-2013
; CURRENT APPLICATION NUMBER: US/10/245,871
; CURRENT FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: 10/197,000
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 905
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 53
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Arachis hypogaea
US-10-245-871-53
Alignment Scores:
Pred. No.: 2,59e-77 Length: 156
Score: 819.00 Matches: 153
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 64.74% Indels: 0
DB: 15 Gaps: 0
US-10-728-051-2 (1-717) x US-10-245-871-53 (1-156)
Qy 2 CTCACTACTAGTAGCCCTCGCCCTTTTCTCTCTCGCTGCCACGCAATCTGGAGGCGAG 61
Db 4 LeuThrIleLeuValAlaLeuAlaLeuPheLeuLeuAlaAlaHisAlaSerAlaArgGln 23
Qy 62 CAGTGGAACTCCAAGGAGACAGAAGATGCCAGAGCCAGCTCAGAGGGCGAACTTGAGG 121
Db 24 GlnTrpGluLeuGlnArgAspArgCysGlnSerGlnLeuGluArgAlaAsnLeuArg 43
Qy 122 CCTGCGAGCAACATCTCATGACAGAGATCCAACTGACGAGGATTCATATGAACGGGAC 181
Db 44 ProCysGluGlnHisLeuMetGlnLysIleGlnArgAspGluAspSerTyrGluArgAsp 63
Qy 182 CCTACAGCCCTAGTACAGGATCCGTACAGCCCTAGTCCATATGATCGAGAGGCGCTTGA 241
Db 64 ProTyrSerProSerGlnAspProTyrSerProSerProTyrAspArgArgGlyAlaGly 83
Qy 242 TCCTCTCAGCACCAGAGAGGTGTGCAATGAGCTGAACAGTTTGAGAACAAACCAAGG 301
Db 84 SerSerGlnHisGlnGluArgCysAsnGluLeuAsnGluPheGluAsnGlnArg 103
Qy 302 TGCATGTGCCAGGCACTTGCACAGATCATCGAGAACCCAGAGCGATAGTTTGCAGGGGAGG 361
Db 124 GlnGlnGluGlnGlnPheLysArgGluLeuArgAsnLeuProGlnGlnCysGlyLeuArg 143
Qy 422 GCACCACAGCGTTGCCGACTTGGACCTGGAACCTCGAAGTGGCGGC 460
Db 144 AlaProGlnArgCysAspLeuAspValGluSerGlyGly 156

Db 104 CysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArgLeuGlnGlyArg 123
Qy 362 CAACAGGAGCAACAGTTCAAGAGGAGCTCAGGAACCTTGCTCAACAGTCCGGCTTAGG 421
Db 124 GlnGlnGluGlnGlnPheLysArgGluLeuArgAsnLeuProGlnGlnCysGlyLeuArg 143
Qy 422 GCACCACAGCGTTGCCGACTTGGACCTGGAACCTCGAAGTGGCGGC 460
Db 144 AlaProGlnArgCysAspLeuAspValGluSerGlyGly 156
RESULT 5
US-10-253-286-53
; Sequence 53, Application US/10253286
; Publication No. US20040058881A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT
; APPLICANT: XU, MINZHEN
; TITLE OF INVENTION: II-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: REH-2015
; CURRENT APPLICATION NUMBER: US/10/253,286
; CURRENT FILING DATE: 2003-01-13
; PRIOR APPLICATION NUMBER: 10/197,000
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 905
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 53
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Arachis hypogaea
US-10-253-286-53
Alignment Scores:
Pred. No.: 2,59e-77 Length: 156
Score: 819.00 Matches: 153
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 64.74% Indels: 0
DB: 15 Gaps: 0
US-10-728-051-2 (1-717) x US-10-253-286-53 (1-156)
Qy 2 CTCACTACTAGTAGCCCTCGCCCTTTTCTCTCTCGCTGCCACGCAATCTGGAGGCGAG 61
Db 4 LeuThrIleLeuValAlaLeuAlaLeuPheLeuLeuAlaAlaHisAlaSerAlaArgGln 23
Qy 62 CAGTGGAACTCCAAGGAGACAGAAGATGCCAGAGCCAGCTCAGAGGGCGAACTTGAGG 121
Db 24 GlnTrpGluLeuGlnArgAspArgCysGlnSerGlnLeuGluArgAlaAsnLeuArg 43
Qy 122 CCTGCGAGCAACATCTCATGACAGAGATCCAACTGACGAGGATTCATATGAACGGGAC 181
Db 44 ProCysGluGlnHisLeuMetGlnLysIleGlnArgAspGluAspSerTyrGluArgAsp 63
Qy 182 CCTACAGCCCTAGTACAGGATCCGTACAGCCCTAGTCCATATGATCGAGAGGCGCTTGA 241
Db 64 ProTyrSerProSerGlnAspProTyrSerProSerProTyrAspArgArgGlyAlaGly 83
Qy 242 TCCTCTCAGCACCAGAGAGGTGTGCAATGAGCTGAACAGTTTGAGAACAAACCAAGG 301
Db 84 SerSerGlnHisGlnGluArgCysAsnGluLeuAsnGluPheGluAsnGlnArg 103
Qy 302 TGCATGTGCCAGGCACTTGCACAGATCATCGAGAACCCAGAGCGATAGTTTGCAGGGGAGG 361
Db 104 CysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArgLeuGlnGlyArg 123
Qy 362 CAACAGGAGCAACAGTTCAAGAGGAGCTCAGGAACCTTGCTCAACAGTCCGGCTTAGG 421
Db 124 GlnGlnGluGlnGlnPheLysArgGluLeuArgAsnLeuProGlnGlnCysGlyLeuArg 143
Qy 422 GCACCACAGCGTTGCCGACTTGGACCTGGAACCTCGAAGTGGCGGC 460
Db 144 AlaProGlnArgCysAspLeuAspValGluSerGlyGly 156

Db 144 AlaProGlnArgCysAspLeuAspValGluSerGlyGly 156

RESULT 6

US-10-100-303A-82
; Sequence 82, Application US/10100303A
; Publication No. US20030202980A1

GENERAL INFORMATION:

; APPLICANT: Caplan, et al.
; TITLE OF INVENTION: Methods and Reagents for Decreasing Clinical Reaction
; FILE OF INVENTION: to Allergy

; FILE REFERENCE: 2002834-0166

; CURRENT APPLICATION NUMBER: US/10/100,303A

; CURRENT FILING DATE: 2002-03-18

; NUMBER OF SEQ ID NOS: 138

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 82

; LENGTH: 166

; TYPE: PRT

; ORGANISM: Arachis hypogaea, Prot/Nucleo Ara h 2

US-10-100-303A-82

Alignment Scores:

Pred. No.:	3.06e-72	Length:	166
Score:	771.00	Matches:	141
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	60.95%	Indels:	0
DB:	15	Gaps:	0

US-10-728-051-2 (1-717) x US-10-100-303A-82 (1-166)

QY 50 TCTCGAGGCGAGCTGGGAACCTCAAGGAGACAGAAGATGCCAGAGCCAGCTCGAGAG 109

Db 15 SerAlaArgGlnGlnTrpGluLeuGlnGlyAspArgCysGlnSerGlnLeuGluArg 34

QY 110 GCGAACCTGAGGCGCTCGGAGCAACATCTCATGCAAGAGATCCAACTGACGAGGATTCA 169

Db 35 AlaAsnLeuArgProCysGluGlnHisLeuMetGlnLysIleGlnArgAspGluAspSer 54

QY 170 TATGAACGGAGCCGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCATATGATCGG 229

Db 55 TyrGluArgAspProTyrSerProSerGlnAspProTyrSerProSerProTyrAspArg 74

QY 230 AGAGGCGCTGATCCTCTGACCAACAGAGAGTGTTCATAGCTGAACGAGTTTGG 289

Db 75 ArgGlyAlaGlySerSerGlnHisGlnGluArgCysAsnGluLeuAsnGluPheGlu 94

QY 290 AACAAACCAAGGTGCATGTGCGAGGCAATTGCAACAGATCATGAGAACAGAGCGATAGG 349

Db 95 AsnAsnGlnArgCysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArg 114

QY 350 TTGAGGGGAGGCAACAGGAGCAACAGTTTCAAGAGGAGCTCAGGAACCTTGCTCAACAG 409

Db 115 LeuGlnGlyArgGlnGlnGluGlnPheLysArgGluLeuArgAsnLeuProGlnGln 134

QY 410 TGGGCGCTTAGGCGACACAGCTTGGAGCTTGGAGCTGCAAGTGGCGGCGAGACAGA 469

Db 135 CysGlyLeuArgAlaProGlnArgCysAspLeuAspValGluSerGlyGlyArgAspArg 154

QY 470 TAC 472

Db 155 Tyr 155

RESULT 7

US-10-899-551-56

; Sequence 56, Application US/10899551

; Publication No. US20050063994A1

GENERAL INFORMATION:

; APPLICANT: Caplan, Michael J.

; APPLICANT: Burke, A. Wesley

; APPLICANT: Sampson, Hugh A.

; APPLICANT: Howard, Sosin B.

; APPLICANT: Bottomly, Kim H.

; TITLE OF INVENTION: Methods and Reagents for Decreasing Clinical Reaction to Allergy
; FILE REFERENCE: 2002834-0233

; CURRENT APPLICATION NUMBER: US/10/899,551

; CURRENT FILING DATE: 2004-07-26

; NUMBER OF SEQ ID NOS: 61

; SOFTWARE: Patentin version 3.2

; SEQ ID NO 56

; LENGTH: 167

; TYPE: PRT

; ORGANISM: species Arachis hypogaea

US-10-899-551-56

Alignment Scores:

Pred. No.:	4.43e-71	Length:	167
Score:	760.00	Matches:	139
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	60.08%	Indels:	0
DB:	17	Gaps:	0

US-10-728-051-2 (1-717) x US-10-899-551-56 (1-167)

QY 53 GCGAGGACAGCTGGGAACCTCAAGGAGACAGAAGATGCCAGAGCCAGCTCGAGAGGCG 112

Db 17 AlaArgGlnGlnTrpGluLeuGlnGlyAspArgCysGlnSerGlnLeuGluArgAla 36

QY 113 AACCTGAGGCGCTCGGAGCAACATCTCATGCAAGAGATCCAACTGACGAGGATTTCATAT 172

Db 37 AsnLeuArgProCysGluGlnHisLeuMetGlnLysIleGlnArgAspGluAspSerTyr 56

QY 173 GAACGGAGCCGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCATATATCGGAGA 232

Db 57 GluArgAspProTyrSerProSerGlnAspProTyrSerProSerProTyrAspArg 76

QY 233 GCGCTGATCCTCTCAGCACCAAGAGAGTGTTCATAGCTGAACGAGTTTGCAGAAC 292

Db 77 GlyAlaGlySerSerGlnHisGlnGluArgCysAsnGluLeuAsnGluPheGluAsn 96

QY 293 AACCAAGGTGCATGTGCGAGGCAATTGCAACAGATCATGAGAACCCAGAGCCATAGTTG 352

Db 97 AsnGlnArgCysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArgLeu 116

QY 353 CAGGGGAGGCAACAGGAGCAACAGTTTCAAGAGGAGCTCAGGAACCTTGCTCAACAGTGC 412

Db 117 GlnGlyArgGlnGlnGluGlnPheLysArgGluLeuArgAsnLeuProGlnGlnCys 136

QY 413 GCGCTTAGGCGACCAAGCGTTCGAGCTTGGAGCTGCAAGTGGCGGCGAGACACAGA 469

Db 137 GlyLeuArgAlaProGlnArgCysAspLeuAspValGluSerGlyGlyArgAspArg 155

RESULT 8

US-09-731-221-77

; Sequence 77, Application US/09731221

; Patent No. US20020018778A1

GENERAL INFORMATION:

; APPLICANT: Caplan, Michael

; TITLE OF INVENTION: Passive Desensitization

; FILE REFERENCE: 2002834-0103

; CURRENT APPLICATION NUMBER: US/09/731,221

; CURRENT FILING DATE: 2001-12-06

; NUMBER OF SEQ ID NOS: 79

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 77

; LENGTH: 166

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence:Arachis

US-09-731-221-77

Alignment Scores:

Pred. No.:	1.21e-64	Length:	166
------------	----------	---------	-----

```
Score: 699.00 Matches: 131
Percent Similarity: 92.91% Conservatives: 0
Best Local Similarity: 92.91% Mismatches: 10
Query Match: 55.26% Indels: 0
DB: 9 Gaps: 0

US-10-728-051-2 (1-717) x US-09-731-221-77 (1-166)

QY 50 TCTGCGAGGAGGAGTGGGAACCTCCAGGAGACAGAGATGCCAGAGCCAGCTCGAGAGG 109
Db 15 SerAlaArgGlnGlnAlaGluLeuGlnGlyAspArgCysGlnSerGlnLeuAlaArg 34
QY 110 GCGAAGCTGAGGCGCTCGGAGCAACATCTCATCAGAGATCCAACTGACGAGGATTCA 169
Db 35 AlaAsnLeuArgAlaCysGluAlaHisLeuMetGlnLysIleGlnAlaAspSer 54
QY 170 TATGAACGGGACCGGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCCATATGATCGG 229
Db 55 TyrGluArgAlaProTyrSerProSerGlnAlaProTyrSerProSerProTyrAspArg 74
QY 230 AGAGCGCTGGATCCTCTCAGCACCAGAGAGGTTGCAATGAGCTGAAACGAGTTTGAG 289
Db 75 ArgGlyAlaGlySerSerGlnHisGlnGluArgCysAsnGluLeuAsnGluPheGlu 94
QY 290 AACAAACCAAGGTGCATGTCGAGGCAATTCACAGATCATGAGACCAAGAGCGATAGG 349
Db 95 AsnAsnGlnArgCysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArg 114
QY 350 TTGCGGGGAGGCAACAGGAGCAACAGTTCAAGAGGAGCTCAGGAACCTGCTCAACAG 409
Db 115 LeuGlnGlyAlaGlnGlnGlnGlnPheGlyArgGluAlaArgAsnLeuProGlnGln 134
QY 410 TCGCGGCTTAGGCACACAGCGTTGCACTTGGAAGTCGAAAGTGGCGGCAGAGACAGA 469
Db 135 CysGlyLeuArgAlaProGlnArgCysAspAlaAspValGluSerGlyGlyArgAspArg 154
QY 470 TAC 472
Db 155 Tyr 155

RESULT 9
US-10-100-303A-81
; Sequence 81, Application US/10100303A
; Publication No. US20030202980A1
; GENERAL INFORMATION:
; APPLICANT: Caplan, et al.
; TITLE OF INVENTION: Methods and Reagents for Decreasing Clinical Reaction
; FILE REFERENCE: 2002834-0166
; CURRENT APPLICATION NUMBER: US/10/100,303A
; CURRENT FILING DATE: 2002-03-18
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 81
; LENGTH: 83
; TYPE: PRT
; ORGANISM: Arachis hypogaea, Prot/Nucleo Ara h 2
US-10-100-303A-81

Alignment Scores:
Pred. No.: 7.94e-40 Length: 83
Score: 463.00 Matches: 83
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 36.60% Indels: 0
DB: 15 Gaps: 0

US-10-728-051-2 (1-717) x US-10-100-303A-81 (1-83)

QY 68 GAATCCCAAGGAGACAGAGATGCCAGAGCTCGAGAGGGCGCAACCTGAGGCCCTGC 127
Db 1 GluLeuGlnGlyAspArgCysGlnSerGlnLeuGluArgAlaAsnLeuArgProCys 20
```

QY	5	ACCATCTAGTACCCCTTGC	CTTTCCTTCCTCGCTGCCACGCATCTCGGAGCGACG	64
Db	5	ThrIleLeuLeuIleSerLeuLeuPheCysIle	--AlaHisThrCysSerAlaSerLys	23
QY	65	TGGGAATCTCAAGGACACAGAAGATGCCAGAGCCAGCTCGAGAGGGCGAACTTCAGGCC	124	
Db	24	TrpGlnHisGlnGlnAsp	--SerCysArgLysGlnLeuGlnGlyAlaLeuLeuThrPro	42
QY	125	TGGGAGCAACATCTCATGCAAGAGATCCAA	-----CCTGACGAGGATTCATATGAACGG	178
Db	43	CysGluLysHisIleMetGluLysIleGlnGlyArgGlyAspAspAspAspAsp	62	
QY	179	GACCCGTACAGCCCTAGTCAGATCCGTACAGCCCTAGTCATATGATTCGGAGAGCGCT	238	
Db	63	AspAspAsnHisIleLeuArgThrMetArgGlyArgIleAsnTrpIleArgAsnGlu	82	
QY	239	GGATCTCTCTCAG	-----CACCAAGAGAGGTGTTCCATATGAGCTGAAC	280
Db	83	GlyLysAspGluAspGluGluGluGluGlyHisMetGlnLysCysThrGluMetSer	102	
QY	281	GAGTTTGAGAACACCAAGGTGCATGTGCGAGGCATTCGCAACAGATCATGGAGAACCG	340	
Db	103	GluLeu	---ArgSerProLysCysGlnCysLysAlaLeuGlnLysIleMetGluAsnGln	121
QY	341	AGCGATAGTTTCAGGGGAGGCAACAGAGCAACAGTTCCTCAAGAGGAGCTCAGGAAC	400	
Db	122	SerGluGluLeuGlu	--GluLysGlnLysLysLysMetGlnLysGluLeuIleAsnLeu	140
QY	401	CCTCAACAGTGGCGCTTAGGCGCACCAAGCGTTGGCACTTG	442	
Db	141	AlaThrMetCysArgPheGlyProMetIleGlnCysAspLeu	154	

QY 125 TCGGAGCAACATCTCATCCAGAAAGTCAA-----CGTGACGAGGATTCAATATGAACGG 178
 Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 62
 43 CysGluLysHisIleMetGluLysIleGlnGlyArgGlyAspAspAspAspAspAsp 62
 QY 179 GACCGGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCATATGATCGGAGAGCGGCT 238
 Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 82
 63 AspAspAsnHisIleLeuArgThrMetArgGlyArgIleAsnIleArgAsnGlu 82
 QY 239 GGATCCTCTCAG-----CACCAGAGAGGTTGTTGCAATGAGCTGAAC 280
 Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 102
 83 GlyLysAspGluAspGluGluGluGlyHisMetGlnCysCysThrGluMetSer 102
 QY 281 GAGTTTGAGAACCAACCAAGGTGCATCTGCAGGCATTGCCAACAGATCATGGAGAACCG 340
 Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 121
 103 GluLeu---ArgSerProLysCysGlnCysLysAlaLeuGlnLysIleMetGluAsnGln 121
 QY 341 AGCCATAGTTGCGAGGGGAGGACACAGGAGCAACAGTTCAGAGGGAGCTCAGGAACCTTG 400
 Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 400
 122 SerGluGluLeuGlu--GluLysGlnLysLysLysMetGluLysGluLeuIleAsnLeu 140
 QY 401 CCTCAACAGTGGCGGCTTAGGGACACACAGCGTTGCGACTTG 442
 Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 154
 141 AlaThrMetCysArgPheGlyProMetIleGlnCysAspLeu 154

RESULT 12

US-10-424-599-264372
 ; Sequence 264372, Application US/10424599
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53223)B
 ; CURRENT APPLICATION NUMBER: US/10/424,599
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 264372
 ; LENGTH: 155
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT3847_80749C.1.pep
 ; US-10-424-599-264372

Alignment Scores:			
Pred. No.:	1.54e-16	Length:	158
Score:	243.00	Matches:	56
Percent Similarity:	56.4%	Conservative:	31
Best Local Similarity:	36.3%	Mismatches:	55
Query Match:	19.21%	Indels:	12
DB:	15	Gaps:	6
US-10-728-051-2 (1-717)	x	US-10-424-599-169225 (1-158)	

```
QY 233 GCGCTGATCCTCTCAG-----CACAAGAGAGGTGTTGCAATGAGCTGAAC 280
Db 80 LysGluGlyLysGluGluGluGluGlyHisMetGlnLysCysSerGluMetSer 99
QY 281 GAGTTTGAGAACCAACAAAGTGTCATGCGGAGGATTGCAACAGATCATGGAGAACCA 340
Db 100 GluLeuLys---SerProIleCysGlnCysLysAlaLeuGlnLysIleMetAspAsnGln 118
QY 341 AGCATAGTTGTCAGGGAGGAGCAACAGAGCAACAGTTTCAAGAGGGAGCTCAGGAATTG 400
Db 119 SerGluGlnLeuGluGlyLys--GluLysLysGlnMetGluArgGluLeuMetAsnLeu 137
QY 401 CCTCAACAGTGGCGCTTAGCGCACCAACAGCGTTGCGACTTG 442
Db 138 AlaIleArgCysArgLeuGlyProMetIleGlyCysAspLeu 151

RESULT 13
US-10-165-289A-2
; Sequence 2, Application US/10165289A
; Publication No. US20030159174A1
; GENERAL INFORMATION:
; APPLICANT: QIU, Xiao
; APPLICANT: TRUSKA, Martin
; APPLICANT: HU, Zhiyuan
; TITLE OF INVENTION: Flax (Linum usitatissimum L.)
; FILE REFERENCE: Seed-Specific Promoters
; CURRENT APPLICATION NUMBER: US/10/165,289A
; PRIOR FILING DATE: 2002-06-06
; PRIOR FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Linum usitatissimum
US-10-165-289A-2

Alignment Scores:
Pred. No.: 1,27e-09 Length: 168
Score: 177.50 Matches: 54
Percent Similarity: 45.71% Conservative: 26
Best Local Similarity: 30.86% Mismatches: 48
Query Match: 14.03% Indels: 47
DB: 14 Gaps: 9

US-10-728-051-2 (1-717) x US-10-165-289A-2 (1-168)
QY 2 CTACCATACTAGTAGCCCTCGCC-----CTTTCTCTCTCGCTGCCACGCACT 52
Db 4 LeuMetSerLeuAlaValAlaThrAlaPheLeuPheLeuIleValValAspAlaSer 23
QY 53 GCGAGGAGCAGTGGGAATCCCAAGGAGAC----- 82
Db 24 ValArgThrThrValIleIleAspGluAspThrAsnGlnGlyArgGlyGlyGly 43
QY 83 -----AGAAGATGCCAGAGCCAGCTC---GAGAGGGCGAACCTGAGGCCCTGC 127
Db 44 GlnGlyGlnGlnGlnCysGluLysGlnIleGlnGlnAspTyrLeuArgSerCys 63
QY 128 GAGCAACATCTCATGACAGATCCACGTCGACAGGATTTCATATGAACGGGCCCTGAC 187
Db 64 GlnGlnPheLeuTrpGluLysValGlnLysGlyArgSerTyr----- 78
QY 188 AGCCCTAGTCAGGATCCGTACAGCCCTAGTCCATCATCGAGAGGCGCTGGA----- 241
Db 79 -----TyrTyrAsnGlnGlyArgGlyGlyGly 87
QY 242 ---TCCTCTCAGCACCAGAGAGGTGTTGCAATGAGCTGAACAGAGTTTGAGAACAA 298
Db 88 GlnGlnSerGlnHisPheAspSerCysCysAspAspLeuLysGlnLeuArgSer----- 105
```

```
QY 299 AGTGCATGTGCGAGGCATTGCAACAGATCATG-----CAGAACCCAGAGCGTAGCTTG 352
Db 106 GluCysThrCysArgGlyLeuGluArgAlaIleGlyGlnMetArgGlnAspIleGlnGln 125
QY 353 CAGCGGAGGCAACAGCAG-----CAACAGTTCAAGAGGAGCTCAGGAAC 397
Db 126 GlnGlyGlnGlnGlnGlnValGluArgTrpValGlnGlnAlaLysGlnValAlaArgAsp 145
QY 398 TTCCCTCAACAGTGGCGCTTAGGGCACCAACAGCGTTGCGACTTG 442
Db 146 LeuProGlyGlnCysGlyThrGln---ProSerArgCysGlnLeu 159

RESULT 14
US-10-382-066-29
; Sequence 29, Application US/10382066
; Publication No. US20040177404A1
; GENERAL INFORMATION:
; APPLICANT: Li, ZhiJian
; APPLICANT: Gray, Dennis
; TITLE OF INVENTION: Nucleotide Sequences of 2S Albumin Gene and its Promoter from Grape
; FILE REFERENCE: Uses Thereof
; CURRENT APPLICATION NUMBER: US/10/382,066
; CURRENT FILING DATE: 2003-03-05
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29
; LENGTH: 167
; TYPE: PRT
; ORGANISM: Vitis vinifera cv. Merlot
US-10-382-066-29

Alignment Scores:
Pred. No.: 1,14e-06 Length: 167
Score: 149.50 Matches: 43
Percent Similarity: 44.51% Conservative: 34
Best Local Similarity: 24.86% Mismatches: 51
Query Match: 11.82% Indels: 45
DB: 16 Gaps: 7

US-10-728-051-2 (1-717) x US-10-382-066-29 (1-167)
QY 2 CTACCATACTAGTAGCCCTCGCCCTTTCTCTCTCGCT---GCCACGCGATCGCGAGG 58
Db 4 LeuSerIlePheAlaIleThrLeuLeuLeuLeuLeuAlaIleSerAsnAlaThrIleTyr 23
QY 59 CAG-----CAGTGGGAACCTCCAAGGA 79
Db 24 GlnThrThrValIleThrArgAspAspGlySerGluPheGlyGlnPheGlnGlySerGln 43
QY 80 GACAGAAGATGCCAGACCCAGCTCGAGAGGGCGAACCTGAGGCCCTCGAGCAACATCTC 139
Db 44 SerGlnArgCysArgGlnGlnIleGlnGlnPheGlnGlnCysGluArgTyrIle 63
QY 140 ATCAGAAGATCCAACTGACGAGGATTTCATATGAACGGGACCCGTACACCCCTAGTCAG 199
Db 64 ArgGlnGlnAlaGluGlnGlnGlnGlyGly----- 73
QY 200 GATCCGTACAGCCCTAGTCCATATGATCGGAGAGGGCGCTGCATCCTCTCAGACCAAGAG 259
Db 74 -----GlnGlyAspValLeuIleLeuArgGlyIleArgAsnGlnGlnGlnGlu 90
QY 260 -----AGGTGTTGCAATGAGCTGAACAGAGTTTGAGAACCAACCAAGGTGC 304
Db 91 GlnGlnTrpLeuArgGlnCysCysGlnAlaLeuGlnAsnMetAsp-----GlnGlnCys 108
QY 305 ATGTCGAGCGCATTCGAACAGATCATGAGAACACAGAGCGCATAGGTGCGAGGG----- 358
Db 109 GlnCysGluGlyLeuArgGlnIleValGlnArgGlnGlnGlyGlnGlyGlnGly 128
QY 359 -----AGGCAACAGGAGCAACAGTTTCAAGAGGAGCTCAGGAAC 397
Db 129 GlnGlyGlnGlyGlnGlyGlnArgGluGlnGlnGlnGluMetMetGlnIleAlaGlnLys 148
```

QY 398 TTGCTCAACAGTGGCGCTTAGGCACACACAGCGTTGC 436
|||||:|||||:|
Db 149 LeuProGluArgCysGly-----SerGlyGlnAlaCys 159
|||:|

RESULT 15

US-10-165-289A-4
; Sequence 4, Application US/10165289A
; Publication No. US20030159174A1
; GENERAL INFORMATION:
; APPLICANT: QIU, Xiao
; APPLICANT: TRUSKA, Martin
; APPLICANT: HU, Zhiyuan
; TITLE OF INVENTION: Flax (Linum usitatissimum L.)
; FILE REFERENCE: Seed-Specific Promoters
; CURRENT APPLICATION NUMBER: US/10/165,289A
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: 60/295823
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Linum usitatissimum
US-10-165-289A-4

Alignment Scores:
Pred. No.: 4.37e-06 Length: 169
Score: 144.00 Matches: 48
Percent Similarity: 45.40% Conservative: 31
Best Local Similarity: 27.59% Mismatches: 47
Query Match: 11.38% Indels: 48
DB: 14 Gaps: 10

US-10-728-051-2 (1-717) x US-10-165-289A-4 (1-169)

QY 2 CTCACCTACTAGTAGCGCTCGCC-----CTTTCTCTCTCGTGCACGCGATCT 52
Db 4 LeuMetSerLeuAlaAlaValAlaThrAlaPheLeuPheLeuValValAspAlaSer 23
|||:|
QY 53 GCGAGGCGACAGTGGGAATC-----73
|||:|
Db 24 ValArgThrThrValIleIleAspGluGluThrAsnGlnGlyArgGlyGlyGlnGly 43
|||:|
QY 74 ---CAAGGACACAGAGA---TGCCAGGCGAGCTCGAGGGCGAAC---CTGAGGCC 124
|||||:|||||:|
Db 44 GlyGlnGlyGlnGlnGlnSerCysGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 63
|||||:|||||:|
QY 125 TCGAGCAACATCTCATGCGAGAGATCCACGTGACGAGGATTCATATGACGGGACCG 184
|||||:|||||:|
Db 64 CysGlnGlnPheMetTrpGluLysValGlnArgGlyGlyArgSerHis----- 79
|||:|
QY 185 TACAGCCCTAGTCAGGATCGGTACAGCCCTAGTCCATATCATCGAGAGCGCTGGA--- 241
|||:|
Db 80 -----TyrThrAsnGlnGlyArgGlyGly 87
|||:|
QY 242 -----TCCTCTGACACAGAGAGGTGTGCAATGAGCTGACGAGTTTGAGAACAA 295
|||||:|||||:|
Db 88 GlyGluGlnSerGlnTyrPheAspSerCysCysAspLeuLysGlnLeuSerThrGly 107
|||||:|||||:|
QY 296 CAAAGGTGCATGTGCGAGGCTTGCACAGATCATG-----CAGAACCGAGCGATAGG 349
|||:|
Db 108 -----CysThrCysArgGlyLeuGluArgAlaIleGlyGlnMetArgGlnGluLeuGln 125
|||:|
QY 350 TTCAGGGGAGGCAACAGAG-----CAACAGTTCAAGAGGGAGCTCAGG 394
|||||:|||||:|
Db 126 GlnGlnGlyGlnGlnGlnGlnValGlnArgTrpIleGlnGlnAlaLysGlnIleAlaLys 145
|||||:|||||:|
QY 395 AACTTCGCTCAACAGTGGCGCTTAGGGACACACAGCGTTGC 436
|||:|
Db 146 AspLeuProGlyGlnCysArgThrGln---ProSerGlnCys 158
|||:|

Search completed: April 8, 2005, 06:48:53
Job time : 85.3048 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model.

Run on: April 8, 2005, 03:43:40 ; Search time 95.8964 Seconds
(without alignments)
5783.472 Million cell updates/sec

Title: US-10-728-051-2

Perfect score: 1265

Sequence: 1 gctccaccatactagtagcc.....taaaagaatcgttttgtt 717

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10728051/runat_07042005_125044_14041/app_query.fasta_1.4757
-DB=A.Geneseq -QFWT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOPECT=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -WATRI=bslosum62 -TRANS=human40.cdi -LIST=45
-DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10728051@cgn_1.1.672 @runat_07042005_125044_14041 -NCPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A.Geneseq_16Dec04.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	842	66.6	157	2 AAW24153	Aaw24153 Peanut al
2	842	66.6	157	2 AAW24164	Aaw24164 Peanut al
3	842	66.6	157	2 AAY15245	Aay15245 Peanut al
4	842	66.6	157	4 AAU04707	Aau04707 Anaphylac
5	842	66.6	157	6 ABUS2464	Abu52464 Peanut Ar
6	842	66.6	157	7 ADG27518	Adg27518 Peanut al
7	838.5	66.3	207	4 AAB82383	Aab82383 Peanut al
8	830	65.6	156	2 AAY40973	Aay40973 Ara h 2 p
9	819	64.7	156	8 ADO38314	Ado38314 Ara h 2 M
10	814	64.3	157	6 ABUS2576	Abu52576 Peanut Ar

11	812	64.2	156	8 ADM12096	Adm12096 Arachis h
12	812	64.2	157	6 ABUS2575	Abu52575 Peanut Ar
13	774	61.2	157	6 ABUS2577	Abu52577 Peanut Ar
14	771	60.9	166	2 AAY40968	Aay40968 Recombina
15	771	60.9	166	6 ABUS2482	Abu52482 Peanut Ar
16	771	60.9	166	7 ADG27536	Adg27536 T7/His-ta
17	699	55.3	166	3 AAB33600	Aab33600 Modified
18	699	55.3	166	4 AAU05035	Aau05035 Modified
19	688.5	54.4	167	4 AAU04710	Aau04710 Modified
20	243	19.2	158	2 AAW23419	Aaw23419 Soybean a
21	243	19.2	158	2 AAY05723	Aay05723 Soybean G
22	243	19.2	158	7 ADH89269	Adh89269 G.max 2S
23	243	19.2	158	8 ADG44004	Adg44004 G.max 2S
24	241.5	19.1	155	2 AAW23418	Aaw23418 Soybean a
25	241.5	19.1	155	7 ADH89267	Adh89267 G.max na
26	241.5	19.1	155	8 ADG44002	Adg44002 G.max 2S
27	229	18.1	158	2 AAW23420	Aaw23420 Chimeric
28	184.5	14.6	323	7 ADH89277	Adh89277 Sunflower
29	184.5	14.6	323	8 ADG44012	Adg44012 H. annuus
30	177.5	14.0	168	6 AAE36076	Aae36076 Flax Conl
31	156	12.3	28	2 AAW24194	Aaw24194 Peanut al
32	156	12.3	158	2 AAW23586	Aaw23586 Mabinlin
33	151	11.9	158	2 AAW23588	Aaw23588 Mabinlin
34	151	11.9	295	7 ADH89275	Adh89275 Sunflower
35	151	11.9	295	8 ADG44010	Adg44010 H. annuus
36	149.5	11.8	167	8 ADS16393	Ads16393 Grape 2S
37	146.5	11.6	164	1 AAP91892	Aap91892 lkb fragm
38	146.5	11.6	164	1 AAP96144	Aap96144 Sequence
39	146.5	11.6	164	7 ADH89227	Adh89227 A. thalia
40	146.5	11.6	164	8 ADG43962	Adg43962 A. thalia
41	144	11.4	169	6 AAE36077	Aae36077 Flax Conl
42	140.5	11.1	146	6 AAO27313	Aao27313 Brazil nu
43	140.5	11.1	146	8 ADN37294	Adn37294 Brazil nu
44	139	11.0	140	1 AAP91891	Aap91891 Brazil nu
45	139	11.0	155	2 AAW23587	Aaw23587 Mabinlin

ALIGNMENTS

RESULT 1

AAW24153
ID AAW24153 standard; protein; 157 AA.

XX AC AAW24153;

XX DT 17-OCT-2003 (revised)

XX DT 29-DEC-1997 (first entry)

XX DE Peanut allergen Ara hII.

XX XX

XX Peanut; seed storage protein; allergen; allergy; hypersensitivity;
KW vaccine; anaphylactic shock; immunotherapy; therapy; monoclonal antibody;
KW ELISA; analysis; Ara hII.

XX OS Arachis hypogaea; strain Florunner.

XX XX

XX WO97241139-A1

XX PD 10-JUL-1997.

XX PF 23-SEP-1996; 96WO-US015222.

XX PR 29-DEC-1995; 95US-0009455P.

XX PR 04-MAR-1996; 96US-00610424.

XX PA (UYAR-) UNIV ARKANSAS.

XX PI Burks AW, Helm RM, Cockrell G, Stanley JS, Bannon GA;

XX DR WPI; 1997-363453/33.

XX DR N-PSDB; AAT76615.

XX PT Peanut allergens Ara hI and Ara hII - used for vaccination and in two-

PT site monoclonal antibody based ELISA.
 XX Claim 31; Page 198; 354pp; English.
 XX This polypeptide comprises major peanut allergen Ara hII. Its sequence
 CC was deduced from a cDNA clone (AA76615) isolated from peanut seed cDNA
 CC using a primer (see AAT76617) based on an isolated Ara hI peptide (see
 CC AAW24151). The sequence shows significant homology with the conglutin
 CC family of seed storage proteins of other legumes. The allergen is
 CC recognised by serum IgE from a large proportion of individuals with
 CC peanut hypersensitivity. Ara hII and Ara hI (see AAW24149-50) can be used
 CC to raise monoclonal antibodies which are used in a specific two-site MAb
 CC ELISA for the detection of Ara hI or Ara hII (claimed). IgE-binding Ara
 CC hII antigen epitopes (see AAW24188-93) may be used in vaccines to protect
 CC against allergic reactions to peanut allergens, e.g. anaphylactic shock.
 CC (Updated on 17-OCT-2003 to standardise OS field)
 XX SQ Sequence 157 AA;
 Alignment Scores:
 Pred. No.: 8-2e-90 Length: 157
 Score: 842.00 Matches: 157
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 66.56% Indels: 0
 DB: 2 Gaps: 0
 US-10-728-051-2 (1-717) x AAW24153 (1-157)
 QY 2 CTCACCATAGTAGCCCTCGCCCTTTCTCTCTCTGCTGCCACGATCTGGAGGCGAG 61
 Db 1 LeuThrIleLeuValAlaLeuAlaLeuPheLeuLeuAlaAlaHisAlaSerAlaArgGln 20
 QY 62 CAGTGGGAAGTCCAGGAGACAGAGATGCCAGAGCCAGCTCGAGAGGCGCACTGAGG 121
 Db 21 GlnTrpGluLeuGlnGlyAspArgGlySerGlnSerGlnLeuGluArgAlaAsnLeuArg 40
 QY 122 CCTGCGAGCAACATCTCTGACAGAGATCCACGTCGACGAGGATTCATATGAACGGGAC 181
 Db 41 ProCysGluGlnHisLeuMetGlnLysIleGlnArgAspGluAspSerTyrGluArgAsp 60
 QY 182 CCGTACAGCCCTAGTCAGATCGGTACAGCCCTAGTCCATATGATCGGAGGCGCTGGA 241
 Db 61 ProTyrSerProSerGlnAspProTyrSerProTyrAspArgGlyAlaGly 80
 QY 242 TCCTCTCAGCACCAGAGAGGTGTGCAATGAGCTGAACGAGTTTGAGAACACCAAGG 301
 Db 81 SerSerGlnHisGlnGluArgGlyCysAsnGlnLeuAsnGluPheGluAsnGlnArg 100
 QY 302 TGCATGTGCGAGGCGATTCGACAGATCATGAGAACCCAGAGCGGATAGGTTGCGGGGAGG 361
 Db 101 CysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArgLeuGlnGlyArg 120
 QY 362 CAACAGAGACACAGTTTCAGAGGAGCTCAGGAACCTGCTCAACAGTCGCGCCCTAGG 421
 Db 121 GlnGlnGluGlnGlnPheLysArgGluLeuArgAsnLeuProGlnGlnCysGlyLeuArg 140
 QY 422 GCACCCAGGCTTGGAGCTGGAGCTGAAAGTGGCGGCGAGACAGATAC 472
 Db 141 AlaProGlnArgCysAspLeuAspValGluSerGlyArgAspArgTyr 157
 RESULT 2
 ID AAW24164
 AC AAW24164 standard; protein; 157 AA.
 XX AAW24164;
 XX 17-OCT-2003 (revised)
 DT 29-DEC-1997 (first entry)
 XX Peanut allergen Ara hII.
 DE Peanut; seed storage protein; allergen; allergy; hypersensitivity;
 .KW

KW vaccine; anaphylactic shock; immunotherapy; therapy; monoclonal antibody;
 KW ELISA; analysis; Ara hII.
 XX Arachis hypogaea; strain Florunner.
 OS WO9724139-A1.
 PN 10-JUL-1997.
 PD 23-SEP-1996; 96WO-US015222.
 PF 29-DEC-1995; 95US-0009455P.
 PR 04-MAR-1996; 96US-00610424.
 XX (UYAR-) UNIV ARKANSAS.
 XX Burks AW, Helm RM, Cockrell G, Stanley JS, Bannon GA;
 XX WPI; 1997-363453/33.
 DR N-PSDB; AAT76614.
 XX Peanut allergens Ara hI and Ara hII - used for vaccination and in two-
 PT site monoclonal antibody based ELISA.
 PT Claim 31; Page 219; 354pp; English.
 PS This polypeptide comprises major peanut allergen Ara hII. Its sequence
 XX was deduced from cDNA clone P38 (AAT76614), isolated from peanut seed
 CC cDNA using a primer (see AAT76617) based on an isolated Ara hI peptide
 CC (see AAW24151). The sequence shows significant homology with the
 CC conglutin family of seed storage proteins of other legumes. The allergen
 CC is recognised by serum IgE from a large proportion of individuals with
 CC peanut hypersensitivity. Ara hII and Ara hI (see AAW24149-50) can be used
 CC to raise monoclonal antibodies which are used in a specific two-site MAb
 CC ELISA for the detection of Ara hI or Ara hII (claimed). IgE-binding Ara
 CC hII antigen epitopes (see AAW24188-93) may be used in vaccines to protect
 CC against allergic reactions to peanut allergens, e.g. anaphylactic shock.
 CC (Updated on 17-OCT-2003 to standardise OS field)
 XX SQ Sequence 157 AA;
 Alignment Scores:
 Pred. No.: 8-2e-90 Length: 157
 Score: 842.00 Matches: 157
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 66.56% Indels: 0
 DB: 2 Gaps: 0
 US-10-728-051-2 (1-717) x AAW24164 (1-157)
 QY 2 CTCACCATAGTAGCCCTCGCCCTTTCTCTCTCTGCTGCCACGATCTGGAGGCGAG 61
 Db 1 LeuThrIleLeuValAlaLeuAlaLeuPheLeuLeuAlaAlaHisAlaSerAlaArgGln 20
 QY 62 CAGTGGGAAGTCCAGGAGACAGAGATGCCAGAGCCAGCTCGAGAGGCGCACTGAGG 121
 Db 21 GlnTrpGluLeuGlnGlyAspArgGlySerGlnSerGlnLeuGluArgAlaAsnLeuArg 40
 QY 122 CCTGCGAGCAACATCTCTGACAGAGATCCACGTCGACGAGGATTCATATGAACGGGAC 181
 Db 41 ProCysGluGlnHisLeuMetGlnLysIleGlnArgAspGluAspSerTyrGluArgAsp 60
 QY 182 CCGTACAGCCCTAGTCAGATCGGTACAGCCCTAGTCCATATGATCGGAGGCGCTGGA 241
 Db 61 ProTyrSerProSerGlnAspProTyrSerProTyrAspArgGlyAlaGly 80
 QY 242 TCCTCTCAGCACCAGAGAGGTGTGCAATGAGCTGAACGAGTTTGAGAACACCAAGG 301
 Db 81 SerSerGlnHisGlnGluArgGlyCysAsnGlnLeuAsnGluPheGluAsnGlnArg 100
 QY 302 TGCATGTGCGAGGCGATTCGACAGATCATGAGAACCCAGAGCGGATAGGTTGCGGGGAGG 361
 Db 101 CysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArgLeuGlnGlyArg 120
 QY 362 CAACAGAGACACAGTTTCAGAGGAGCTCAGGAACCTGCTCAACAGTCGCGCCCTAGG 421
 Db 121 GlnGlnGluGlnGlnPheLysArgGluLeuArgAsnLeuProGlnGlnCysGlyLeuArg 140
 QY 422 GCACCCAGGCTTGGAGCTGGAGCTGAAAGTGGCGGCGAGACAGATAC 472
 Db 141 AlaProGlnArgCysAspLeuAspValGluSerGlyArgAspArgTyr 157
 RESULT 2
 ID AAW24164
 AC AAW24164 standard; protein; 157 AA.
 XX AAW24164;
 XX 17-OCT-2003 (revised)
 DT 29-DEC-1997 (first entry)
 XX Peanut allergen Ara hII.
 DE Peanut; seed storage protein; allergen; allergy; hypersensitivity;
 .KW

Db 101 CysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArgLeuGlnGlyArg 120
 QY 362 CAACAGAGCAACAGTTCAAGAGGGAGCTCAGGAAGTTGCTCAACAGATCGGCCTTAGG 421
 Db 121 GlnGlnGluGlnGlnPheLysArgGluLeuArgAsnLeuProGlnGlnCysGlyLeuArg 140
 QY 422 GCACCAAGCGTTGCGACTTGGAGCTCGAAAGTGGCGGAGAGACAGATAC 472
 Db 141 AlaProGlnArgCysAspLeuAspValGluSerGlyGlyArgAspArgTyr 157

RESULT 3
 AAY15245
 ID AAY15245 standard; protein; 157 AA.
 XX AC AAY15245;
 XX DT 17-OCT-2003 (revised)
 XX DT 09-NOV-1999 (first entry)
 DE Peanut allergen, Ara h 2, amino acid sequence.
 XX KW allergy; immune response; transgenic; allergen; epitope;
 KW immunoglobulin E; Ig E; binding site; peanut.
 XX OS Arachis hypogaea.
 XX PN WO9938978-A1.
 XX PD 05-AUG-1999.
 XX PF 29-JAN-1999; 99WO-US002031.
 XX PR 31-JAN-1998; 98US-0073283P.
 PR 13-FEB-1998; 98US-0074590P.
 PR 13-FEB-1998; 98US-0074624P.
 PR 13-FEB-1998; 98US-0074633P.
 PR 27-AUG-1998; 98US-00141220.
 XX (UYAR-) UNIV ARKANSAS.
 PA (UYNY) UNIV NEW YORK MT SINAI SCHOOL MEDICINE.
 PA (SOSI/) SOSIN H.
 XX PI Sosin H, Bannon GA, Burks AW, Sampson HA;
 XX WPI; 1999-479189/40.
 DR N-PSDB; AA206383.
 XX PT Modified allergen with reduced IgE binding, useful for treating e.g.
 PT allergies.
 XX PS Disclosure; Page 38; 46pp; English.
 CC This is the amino acid sequence of the Ara h 2 protein from Arachis
 CC hypogaea. The Ara h 2 protein has 10 IgE (immunoglobulin E) binding
 CC epitopes, three of which are immunodominant (AAY15272, AAY15275, and
 CC AAY15276). By modifying the IgE binding sites the ability of the allergen
 CC to provoke an immune response is downregulated. The epitopes of the IgE
 CC binding sites can therefore be modified in genetically engineered plants
 CC and animals to elicit less of an allergic response. (Updated on 17-OCT-
 CC 2003 to standardise OS field)
 XX SQ Sequence 157 AA;

Alignment Scores:
 Pred. No.: 8, 2e-90 Length: 157
 Score: 842.00 Matches: 157
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 66.56% Indels: 0
 DB: 2 Gaps: 0

US-10-728-051-2 (1-717) x AAY15245 (1-157)

QY 2 CTCACACTACTAGTACCCCTCGCCCTTTTCCTCCTCGCTGCCACCGCATCTCGAGGACG 61
 Db 1 LeuThrIleLeuValAlaLeuAlaLeuPheLeuLeuAlaHisAlaSerAlaArgGln 20
 QY 62 CAGTGGAACTCCCAAGGAGACAGAGTCCAGAGGCGAGCTCGAGAGGGGGAACCTGAGG 121
 Db 21 GlnTrpGluLeuGlnGlyAspArgCysGlnSerGlnLeuGluArgAlaAsnLeuArg 40
 QY 122 CCTCGAGCAACATCTCATGCAAGAGATCCCAACGTGACGAGGATTTCATATGAACGGAC 181
 Db 41 ProCysGluGlnHisLeuMetGlnLysIleGlnArgAspGluAspSerTyrGluArgAsp 60
 QY 182 CGGTACAGCCCTAGTACGATCCGTACAGCCCTAGTCCATATCATCGGAGAGGGCTGGA 241
 Db 61 ProTyrSerProSerGlnAspProTyrSerProSerProTyrSerProTyrSerProTyr 80
 QY 242 TCCTCTCAGCACCAAGAGAGGTGTTGCAATGAGCTCAACGAGTTTTCAGAACCAACAAAG 301
 Db 81 SerSerGlnHisGlnGluArgCysCysAsnGluLeuAsnGluPheGluAsnGlnArg 100
 QY 302 TGCATGTGCGAGGCATTGCAACAGATCATGGAGAACCCAGAGCGATAGTTTCAGGGGAGG 361
 Db 101 CysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArgLeuGlnGlyArg 120
 QY 362 CAACAGAGCAACAGTTCAAGAGGGAGCTCAGGAAGTTGCTCAACAGATCGGCCTTAGG 421
 Db 121 GlnGlnGluGlnGlnPheLysArgGluLeuArgAsnLeuProGlnGlnCysGlyLeuArg 140
 QY 422 GCACCAAGCGTTGCGACTTGGAGCTCGAAAGTGGCGGAGAGACAGATAC 472
 Db 141 AlaProGlnArgCysAspLeuAspValGluSerGlyGlyArgAspArgTyr 157

RESULT 4
 AAU04707
 ID AAU04707 standard; protein; 157 AA.
 XX AC AAU04707;
 XX DT 23-OCT-2001 (first entry)
 XX DE Anaphylactic antigen Ara h 2.
 XX KW Ara h 2; anaphylactic antigen; immunoglobulin E; IgE; immunogenic;
 KW allergy; mast cell; basophil; mouse.
 XX OS Mus sp.
 XX PN WO200140264-A2.
 XX PD 07-JUN-2001.
 XX PF 06-DEC-2000; 2000WO-US033124.
 XX PR 06-DEC-1999; 99US-00455294.
 PR 23-JUN-2000; 2000US-0213765P.
 PR 27-SEP-2000; 2000US-0235797P.
 XX (PANA-) PANACEA PHARM LLC.
 PA (UYAR-) UNIV ARKANSAS.
 PA (MOUN) MOUNT SINAI SCHOOL MEDICINE NEW YORK.
 XX Bannon GA, Burks WA, Caplan MJ, Sampson H, Sosin H;
 XX WPI; 2001-381378/40.
 XX PT Antigenic fragments useful for reducing anaphylactic risk and reducing
 PT the severity and/or number of allergic symptoms in individuals sensitive
 PT to antigens, have reduced ability to bind immunoglobulin E.
 XX Claim 7; Fig 10; 100pp; English.
 XX The sequence represents the amino acid sequence of anaphylactic antigen
 CC Ara h 2. Ara h 2 is an anaphylactic antigen (A), which was used to design

antigenic peptides having a reduced ability to bind immunoglobulin E (IgE) as compared with the intact (A), or having a sequence substantially identical to a portion of sequence of an antigen that includes at least one IgE binding site, where at least one IgE binding site of the peptide is altered. The antigenic peptides are used in a composition which is useful for reducing risk or severity of allergic reaction to an antigen. This is done by identifying an individual at risk of allergic reaction to an antigen by identifying prior display of allergic symptoms when exposed to the antigen, or a familial relationship with an individual who previously displayed allergic symptoms when exposed to the antigen. Following this an antigen-specific IgE present on one or more mast cells or basophils in the individual's serum is identified. The individual is then contacted with a peptide corresponding to a portion of the antigen, which is selected, formulated, and delivered so that binding of the peptide to antigen-specific IgE is reduced as compared with IgE binding of intact antigen. The composition is also useful for treating and preventing allergic reactions

Sequence 157 AA;

Alignment Scores:		
Pred. No.:	8.2e-90	Length: 157
Score:	842.00	Matches: 157
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	66.56%	Indels: 0
DB:	4	Gaps: 0

US-10-728-051-2 (1-717) x AAU04707 (1-157)

QY	2	CTCACCATACTAGTAGCCCTCGCTTCTCTCGCTGCCACGCAATCTGCAGGCAG	61
Db	1	LeuThrIleLeuValAlaLeuAlaLeuPheLeuLeuAlaHisAlaSerAlaAraGln	20
QY	62	CAGTGGGAACCTCAAGGAGACAGAAGATGCCAGAGCCAGCTCGAGAGGCGCAACCTGAGG	121
Db	21	GlnTrpGluLeuGlnGlyAspArgCysGlnSerGlnLeuGluAraAlaAsnLeuAraG	40
QY	122	CCCTGGAGCAACATCTCATGACAGAAGATCCAACTGACGAGGAGNTTCATATGAACCGGAC	181
Db	41	ProCysGluGlnHisLeuMetGlnIleGlnArgAspGluAspSerTyrGluAraGAsp	60
QY	182	CGGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCCATATGATTCGAGAGCGCTGGA	241
Db	61	ProTyrSerProSerGlnAspProTyrSerProSerProTyrAspArgGlyAlaGly	80
QY	242	TCCTCTCAGCACCAAGAGAGGTGTTGCAATCAGCTGAACGAGTTTGAGAACCAACCAAGG	301
Db	81	SerSerGlnHisGlnGluArgCysCysAsnGluLeuAsnGluPheGluAsnAsnGlnAraG	100
QY	302	TGCATGTGCGAGGCATTGCCACAGATCATGAGAACCCAGAGCCAGTAGTTGCAGGGGAGG	361
Db	101	CysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArgLeuGlnGlyAraG	120
QY	362	CAACAGGAGCAACAGCTTCAAGAGGAGGAGCTCAGGAACCTTGCTCAACAGTCGGCGCTTAGG	421
Db	121	GlnGlnGluGlnGlnPheIleYaraGluLeuAraGlnLeuProGlnGlnCysGlyLeuAraG	140
QY	422	GCACCAACAGCGTTGCGATTGGACGTGAAAGTGGCGGCAGACAGATAC	472
Db	141	AlaProGlnAraCysAspLeuAspValGluSerGlyGlyAraAspAraTyr	157

RESULT 5
ABU52464

ABU52464
ID ABU52464 standard; protein; 157 AA.

AC ABU52464;

DT 10-MAR-2003 (first entry)

DE Peanut Ara h2 protein.

KW Peanut; allergy; Ara h1; Ara h2; Ara h3; IgE binding site;

٢٠

KW	anaphylactic food allergen; antiallergenic; vaccine; wound healing.
XX	
OS	Arachis hypogaea.
XX	
FN	WO200274250-A2.
XX	
XX	26-SEP-2002.
PD	
XX	
XX	18-MAR-2002; 2002WO-US009108.
PF	
XX	16-MAR-2001; 2001US-0276822P.
PR	18-MAR-2002; 2002US-00276822.
PR	
XX	
XX	(PANA-) PANACEA PHARM.
PA	
XX	
XX	Caplan M, Sosin H, Sampson H, Bannon GA, Burks WA, Cockrell G;
PI	Compadre CM, Connaughton C, Helm NE, King NE, Kopper RA, Maleki SJ;
PI	Rabjohn PA, Shin DS, Stanley JS;
XX	
XX	WPI: 2003-018765/01.
DR	
DR	N-PSDB; ABX70606.
XX	
XX	
PT	New modified anaphylactic food allergen, useful for preventing or
PT	treating allergic reactions associated with e.g. anaphylactic allergens.
XX	
PS	Claim 27; Fig 41: 300bp; English.

Alignment Scores:

Pred. No.:	8.2e-90	Length:	157
Score:	842.00	Matches:	157
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	66.56%	Indels:	0
DB:	6	Gaps:	0

US-10-728-051-2 (1-717) x ABU52464 (1-157)

QY	2	CTCACATACATAGTAGCCCTCGCCCTTTCTCTCTCGCTGCCACGCAATCTCGAGGCAG	61
Db	1	LeuThrIleLeuValAlaLeuAlaLeuPheLeuLeuAlaAlaHisAlaSerAlaArgGln	20
QY	62	CAGTGGGAATCCAGGAGACAGAGATGCCAGAGCCAGCTCAGAGGGCGGCACTTCAGG	121
Db	21	GlnTrpGluLeuGingIlyAspArgCysGlnSerGlnLeuGluArgAlaAsnLeuArg	40
QY	122	CCTCGCAGCAACATCTCATCCAGAAGATCCAAAGTCACGAGGATTTCATATGAACGGGAC	181
Db	41	ProCysGluGlnHisLeuMetGlnIlyIleGlnArgAspGluAspSerTyrGluArgAsp	60
QY	182	CCGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCCATATCATCGAGAGCGCTGGA	241
Db	61	ProTyrSerProSerGlnAspProTyrSerProSerProTyrAspArgAGlValaGly	80

```
QY 242 TCCTCTCAGCACCAAGAGAGGTTTGCATGAGCTGAAACAGTTTGAGAACAAACGAGG 301
DB 81 SerSerGlnHisGlnGluArgCysCysAsnGluLeuAsnGluPheGluAsnAsnGlnArg 100
QY 302 TGCATGTGCGAGGCACTTGCACACAGATCATGAGAACCCAGAGCGATAGTTGCAGGGGAGG 361
DB 101 CysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArgLeuGlnGlyArg 120
QY 362 CAACAGAGCACACAGTTCAAGAGGGAGCTCAGGAACCTTGCTCAACAGATGCGGCCTTAGG 421
DB 121 GlnGlnGluGlnGlnPheLysArgGluLeuArgAsnLeuProGlnGlnCysGlyLeuArg 140
QY 422 GCACACAGGCTTGCAGCTTGGACGTGCGAAAGTGGCGGCGAGACAGATAC 472
DB 141 AlaProGlnArgCysAspLeuAspValGluSerGlyGlyArgAspArgTyr 157

RESULT 6
ADG27518
ID ADG27518 standard; protein; 157 AA.
XX
AC ADG27518;
DT 26-FEB-2004 (first entry)
XX
DE Peanut allergen Ara h2.
XX
KW Peanut; plant; allergen; Ara h1; Ara h2; Ara h3; glycinin A2B1a; Jug n1;
KW antiallergic; vullnerary; anaphylactic food allergen; IGE; allergy; wound.
XX
OS Arachis hypogaea.
XX
US2003202980-A1.
XX
30-OCT-2003.
XX
18-MAR-2002; 2002US-00100303.
XX
29-DEC-1995; 95US-0009455P.
PR 23-SEP-1996; 96US-00717933.
PR 31-JAN-1998; 98US-0073283P.
PR 13-FEB-1998; 98US-0074590P.
PR 13-FEB-1998; 98US-0074624P.
PR 13-FEB-1998; 98US-0074633P.
PR 29-JUN-1998; 98US-00106872.
PR 27-AUG-1998; 98US-00141220.
PR 13-NOV-1998; 98US-00191593.
PR 29-JAN-1999; 99US-00240557.
PR 29-JAN-1999; 99US-00241101.
PR 11-FEB-1999; 99US-00248673.
PR 11-FEB-1999; 99US-00248674.
PR 02-MAR-1999; 99US-0122450P.
PR 02-MAR-1999; 99US-0122452P.
PR 02-MAR-1999; 99US-0122565P.
PR 02-MAR-1999; 99US-0122566P.
PR 11-MAR-1999; 99US-00267719.
PR 28-JAN-2000; 2000US-00494096.
PR 16-MAR-2001; 2001US-0276822P.
XX
(CAPL/) CAPLAN M J.
PA (SOSI/) SOSIN H B.
PA (SAMP/) SAMPSON H.
PA (BANN/) BANNON G A.
PA (BURK/) BURKS A W.
PA (COCK/) COCKRELL G.
PA (COMP/) COMPADRE C M.
PA (CONN/) CONNAUGHTON C.
PA (HELM/) HELM R M.
PA (KING/) KING N E.
PA (KOPP/) KOPPER R A.
PA (MALE/) MALEKI S J.
PA (RABJ/) RABJOHN P A.
```

```
PA (SHIN/) SHIN D S.
PA (STAN/) STANLEY J S.
PI Caplan MJ, Sosin HB, Sampson H, Bannon GA, Burks AW, Cockrell G,
PI Compadre CM, Connaughton C, Helm RM, King NE, Kopper RA, Maleki SJ,
PI Rabjohn PA, Shin DS, Stanley JS;
XX WPI; 2003-875632/81.
DR N-PSDB; ADG27517.
XX
XX New modified anaphylactic food allergen comprising a cysteine residue
XX which has been modified so that it cannot participate in the disulfide
XX bond, useful for treating allergic reactions or wounds.
XX
XX Claim 27; SEQ ID NO 63; 194pp; English.
XX
XX The invention relates to a modified anaphylactic food allergen whose
XX amino acid sequence is substantially identical to that of a natural
XX anaphylactic food allergen. The natural anaphylactic food allergen
XX includes at least one cysteine residue that participates in a disulphide
XX bond when the natural anaphylactic food allergen is in its native
XX conformation, except that the cysteine residue has been modified so that
XX it cannot participate in the disulphide bond. Also included are a method
XX of making a modified anaphylactic food allergen, a nucleotide molecule
XX encoding a modified anaphylactic food allergen defined above, a
XX nucleotide molecule for causing a site specific mutation in a gene
XX encoding a natural anaphylactic food allergen, a transgenic plant or
XX animal expressing a modified anaphylactic food allergen defined above, a
XX method of treating an individual by reducing the clinical response to a
XX natural anaphylactic food allergen by administering a modified
XX anaphylactic food allergen and an isolated fragment of peanut allergen
XX Ara h 1, comprising at least 10 consecutive amino acids of ADG27464 or
XX ADG27465. About 10-17% of the amino acids have been modified in at least
XX one IGE epitope or all the IGE epitopes recognised when the natural
XX anaphylactic food allergen is contacted with serum IGE from individual(s)
XX allergic to the natural anaphylactic food allergen. The invention
XX discloses Peanut allergens Ara h1, Ara h2, Ara h3 (and their encoding
XX cDNAs), Soybean Glycinin A2B1a and IGE-binding epitopes of the English
XX walnut allergen Jug n1. The modified anaphylactic food allergen can be
XX used for treating allergic reactions or wounds. The present sequence
XX represents a Peanut allergen of the invention (or its fragment).
XX
XX Sequence 157 AA;
```

```
Alignment Scores:
Pred. No.: 8.2e-90 Length: 157
Score: 842.00 Matches: 157
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 66.56% Indels: 0
DB: 7 Gaps: 0
```

US-10-728-051-2 (1-717) x ADG27518 (1-157)

```
QY 2 CTCACCATACTAGTACCCCTCGCCCTTTCCTCTCCTCGCTGCCCCAGCATCTCGAGGCAG 61
DB 1 LeuThrIleLeuValAlaLeuAlaLeuPheLeuLeuAlaHisAlaSerAlaArgGln 20
QY 62 CAGTGGGAAGTCCCAAGGAGACAGAAAGTGCAGAGCCAGCTCGAGAGGGGGAACCTGAGG 121
DB 21 GlnTrpGluLeuGlnGlyAspArgAArgCysGlnSerGlnLeuGluArgAlaAsnLeuArg 40
QY 122 CCCTGCGAGCACATCTCATGCGAAGATCCAAAGTGCAGAGGATTCATATGAACGGGAC 181
DB 41 ProCysGluGlnHisLeuMetGlnLysIleGlnArgAspGluAspSerTyrGluArgAsp 60
QY 182 CCGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCCATATCATCGGAGAGGGCTGGA 241
DB 61 ProTyrSerProSerGlnAspProTyrSerProSerProTyrAspArgArgGlyAlaGly 80
QY 242 TCCTCTCAGCACCAAGAGAGGTTTGCATGAGCTGAAACAGTTTGAGAACAAACGAGG 301
DB 81 SerSerGlnHisGlnGluArgCysCysAsnGluLeuAsnGluPheGluAsnAsnGlnArg 100
```

QY 302 TGCATGTGCGAGGATTGCAACAGATCATCGAGAACCCAGAGCCGATAGTTGCGAGGGAGG 361
 DB 101 CysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArgLeuGlnGlyArg 120
 QY 362 CAACAGGAGCAACAGTTCAAGAGGGAGCTCAGGAAGTTCCTCAACAGTCGGGCTTAGG 421
 DB 121 GlnGlnGlnGlnGlnPhelYsArgGluLeuArgAsnLeuProGlnGlnCysGlyLeuArg 140
 QY 422 GCACCAAGCGTTGCGACTTGGACGTCGAAAGTGGCGGCGAGACAGATAC 472
 DB 141 AlaProGlnArgCysAspLeuAspValGluSerGlyGlyArgAspArgTyr 157

RESULT 7
 AAB82383
 ID AAB82383 standard; protein; 207 AA.
 XX
 AC AAB82383;
 XX
 DT 11-SEP-2003 (revised)
 DT 23-JUL-2001 (first entry)
 XX
 DE Peanut allergen Ara h2 gene product.
 XX
 KW Peanut; allergen; Ara h2; transgenic plant; allergy.
 XX
 OS Arachis hypogaea.
 XX
 FH Location/Qualifiers
 FT Peptide 1..21
 FT Protein /label= Signal_peptide
 FT /label= Mature_protein
 XX
 PN WO200136621-A2.
 XX
 XX 25-MAY-2001.
 XX
 PF 20-NOV-2000; 2000WO-US031657.
 XX
 PR 19-NOV-1999; 99US-0167255P.
 XX
 PA (UYAL-) UNIV ALABAMA A & M.
 XX
 PI Dodo HW, Arntzen CJ, Konan KN, Viquez OM;
 XX
 DR WPI; 2001-355630/37.
 DR N-PSDB; AAF90336, AAF90337.
 XX
 PT Producing transgenic peanut plants that produce allergen-free seeds,
 PT useful in non-allergenic foods, by antisense or sense co-suppression of
 PT allergen-encoding genes.
 XX
 XX Example 1; Fig 2; 72pp; English.
 XX
 CC The present sequence is that of the peanut allergenic protein (AP)
 CC encoded by the Ara h2 gene (see AAF90336). The invention relates to a
 CC method for producing a peanut plant having reduced, or undetectable, AP
 CC content in its seed. A peanut plant cell is transformed with a DNA
 CC construct containing an antisense AP gene and/or sense AP gene, or their
 CC fragments, regenerated to plants, and fertile transgenic plants that
 CC produce seeds with reduced AP content are identified. The AP sense or
 CC antisense gene may comprise at least a portion of the Ara h2 gene
 CC sequence. The seeds are useful for preparation of allergen-free foods.
 CC Recombinant AP may be produced and used to produce antibodies useful for
 CC detecting AP in foods, and for treatment or prevention of peanut allergy.
 CC (Updated on 11-SEP-2003 to standardise OS field)
 XX
 SQ Sequence 207 AA;

Alignment Scores: 2.39e-89 Length: 207
 Pred. No.: 838.50 Matches: 162
 Score:

Percent Similarity: 90.27% Conservative: 5
 Best Local Similarity: 87.57% Mismatches: 17
 Query Match: 66.28% Indels: 1
 DB: 4 Gaps: 1

US-10-728-051-2 (1-717) x AAB82383 (1-207)

QY 2 CTCACCATAGTAGCCCTCGCCCTTTTCCTCGCTGCCACGCATCTGCGAGGCAG 61
 DB 4 LeuThrIleLeuValAlaLeuAlaLeuPheLeuLeuAlaHlaHlaSerAlaArgGln 23
 QY 62 CAGTGGGAATCTCAAGGAGACAGATGCCAGAGCCAGCTCGAGAGGCGCAACCTGAG 121
 DB 24 GlnTrpGluLeuGlnGlyAspArgCysGlnSerGlnLeuGluArgAlaAsnLeuArg 43
 QY 122 CCTCGGAGCAACATCTCATGCGAGAGATCCAACTGAGAGGATTCATATCAACGGGAC 181
 DB 44 ProCysGluGlnHisLeuMetGlnIleGlnArgAspGluAspSerTyrGluArgAsp 63
 QY 182 CCGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCCATATATCGGAGAGCGCTGGA 241
 DB 64 ProTyrSerProSerGlnAspProTyrSerProSerProTyrAspArgGlyAlaGly 83
 QY 242 TCCTCTCAGCACCAAGAGAGGTGTTGCAATGAGCTGACGAGTTTGAGAACCAACGAG 301
 DB 84 SerSerGlnHisGlnGluArgCysCysAsnGluLeuAsnGluPheGluAsnGlnArg 103
 QY 302 TGCATGTGCGAGGATTGCAACAGATCATGGAGAACCCAGAGCGATAGTTGCGAGGGAGG 361
 DB 104 CysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArgLeuGlnGlyArg 123
 QY 362 CAACAGGAGCAACAGTTCAAGAGGGAGCTCAGGAATTCCTCAACAGTCGGGCTTAGG 421
 DB 124 GlnGlnGlnGlnGlnPheLysArgGluLeuArgAsnLeuProGlnGlnCysGlyLeuArg 143
 QY 422 GCACCAAGCGTTGCGACTTGGACGTCGAAAGTGGCGGCGAGACAGATACATAACACCT 481
 DB 144 AlaProGlnArgCysAspLeuAspValGluSerGlyGlyArgProArgIleProPro 163
 QY 482 ATCTCAAAAAAGAAAAAGAAAAAGAAATAGCTTATATATAAGC---TATTATCTA 538
 DB 164 IleLeuThrGlySerArgSerArgArgHisGlnSerProTyrGlyAsnArgArgTyrSer 183
 QY 539 TGGTTATGTTAGTT 553
 DB 184 AlaMetCysLeuLeu 188

RESULT 8
 AAY40973
 ID AAY40973 standard; protein; 156 AA.
 XX
 AC AAY40973;
 XX
 DT 17-OCT-2003 (revised)
 DT 06-DEC-1999 (first entry)
 XX
 DE Ara h 2 protein fragment.
 XX
 KW Peanut; allergen; Ara h 1; IgE; immunoglobulin E; epitope; Ara h 3;
 KW allergic reaction; Ara h 2.
 XX
 OS Arachis hypogaea.
 XX
 PN WO9945961-A1.
 XX
 PD 16-SEP-1999.
 XX
 PF 12-MAR-1999; 99WO-US005494.
 XX
 PR 12-MAR-1998; 98US-0077763P.
 PR 11-MAR-1999; 99US-0007763.
 XX
 PA (UYAR-) UNIV ARKANSAS.

XX
PI Burks W, Helm RM, Cockrell G, Bannon GA, Stanley JS, Shin DS;
PI Sampson H, Compadre CM, Huang SK, Maleki SU, Kopper RA;
XX WPI; 1999-551218/46.
XX
PT Tertiary structure of peanut allergen Ara h 1 for protection of a host
PT animal from allergic reaction.
XX
PS Disclosure; Page 104; 193pp; English.
XX
XX The invention provides a tertiary structure for the peanut allergen Ara H
CC 1. The Ara H 1 allergen is found to contain 23 linear IgE-binding
CC epitopes. The invention also provides an isolated recombinant peanut
CC allergen designated Ara h 3 and a nucleotide molecule encoding the peanut
CC allergen Ara h 3. Molecules of the invention are used to protect a host
CC animal from allergic reaction, particularly using a modified allergen
CC which is less reactive with IgE. The invention may also be used to ensure
CC that the allergen is not introduced into genetically modified food. The
CC present sequence represents an Ara h 2 protein fragment. (Updated on 17-
CC OCT-2003 to standardise OS field)
XX
SQ Sequence 156 AA;

Alignment Scores:
Pred. No.: 2.13e-88 Length: 156
Score: 830.00 Matches: 155
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 65.61% Indels: 0
DB: 2 Gaps: 0

US-10-728-051-2 (1-717) x AAY40973 (1-156)
QY 2 CTCACTACTAGTAGCCCTCGCCCTTTCTCTCTCGTCCACGCGATCGGAGCGAG 61
Db 1 LeuThrIleuValalaLeuAlaLeuPheLeuLeuAlaAlaHisalaSerAlaArgGln 20

QY 62 CAGTGGGAATCCACGAGACAGAGATGCCAGAGCCAGCTCAGAGGGCGCACTGAGG 121
Db 21 GlnTrpGluLeuGlnGlyAspArgCysGlnSerGlnLeuGluArgAlaAsnLeuArg 40

QY 122 CCCTGGGAGCAATCTCATGCGAGAGATCCACGTCGACAGGATTCATATGACGGGAC 181
Db 41 ProCysGluGlnHisLeuMetGlnLysIleGlnArgAspGluAspSerTyrGluArgAsp 60

QY 182 CCGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCCATATGATCGAGAGCGCTGGA 241
Db 61 ProTyrSerProSerGlnAspProTyrSerProSerProTyrAspArgGlyAlaGly 80

QY 242 TCCTCTCAGCACCAAGAGAGGTGTGCAATGAGCTGAACGAGTTTGAGAACCAAGG 301
Db 81 SerSerGlnHisGlnGluArgCysAsnGluLeuAsnGluPheGluAsnAsnGlnArg 100

QY 302 TGATGTGCGAGGATTTGCAACAGATCATGAGACACAGAGCGATAGTTGCGAGGAGG 361
Db 101 CysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArgLeuGlnGlyArg 120

QY 362 CAACAGGAGCAACAGTTCAAGAGGGAGCTCAGGAACCTTGCTCAACAGTCGGCCTTAGG 421
Db 121 GlnGlnGlnGlnPheLysArgGluLeuArgAsnLeuProGlnGlnCysGlyLeuArg 140

QY 422 GCACACAGGCTTGCGACTTGGACGTGCGAAGTGGCGGCGAGAC 466
Db 141 AlaProGlnArgCysAspLeuAspValGluSerGlyGlyArgAsp 155

RESULT 9
ADO38314
ID ADO38314 standard; peptide; 156 AA.
XX
AC ADO38314;
XX
DT 15-JUL-2004 (first entry)

XX
DE Ara h 2 MHC-class II-presented epitope #1.
XX
KW Antibacterial; Virucide; Fungicide; Antiparasitic; Antiarthritic;
KW Antirheumatic; Neuroprotective; Antiinflammatory; Dermatological;
KW Immunosuppressive; Antidiabetic; Antichyroid; Antiasthmatic;
KW Antiallergic; Cytostatic; Antipsoriatic; Gene Therapy; Vaccine;
KW MHC Class II; Ii-Key motif; immune response; anthrax; EBOLA; HIV;
KW Influenza; vaccinia virus; infection; bacterium; virus; parasite; fungus;
KW rickettsia; rheumatoid arthritis; multiple sclerosis;
KW lupus erythematosus; diabetes mellitus; myasthenia gravis;
KW autoimmune thyroiditis; scleroderma; dermatomyositis; pemphigus; asthma;
KW allergic rhinitis; topical dermatitis; colitis; cancer; psoriasis;
KW adenoma; peanut; Ara h 2.
XX
OS Arachis hypogaea.
XX
PN US2004058881-A1.
XX
PD 25-MAR-2004.
XX
PF 24-SEP-2002; 2002US-00253286.
XX
PR 24-SEP-2002; 2002US-00253286.
XX
PA (ANTI-) ANTIGEN EXPRESS INC.
XX
PI Humphreys RE, Xu M;
XX
DR WPI; 2004-294259/27.
XX
PT New non-naturally occurring protein or polypeptide modified by
PT recombinant DNA techniques, useful for treating multiple sclerosis,
PT diabetes mellitus, myasthenia gravis, scleroderma, allergic rhinitis,
XX colitis, cancer or psoriasis.
PS Example 2; Page 18; 90pp; English.

CC The invention relates to a non-naturally occurring protein or polypeptide
CC (I) modified by recombinant DNA techniques comprising: a C-terminal
CC element comprising an MHC Class II-presented epitope; an N-terminal
CC element comprising an Ii-key motif; and an intervening element comprising
CC a sequence of 4-11 amino acid residues where the modification by
CC recombinant DNA techniques taking place within elements (b) and (c). Also
CC described are methods for: suppressing or enhancing an immune response
CC directed toward an MHC (major histocompatibility complex) Class II-
CC presented epitope of interest. Suppressing an immune response directed
CC toward an MHC Class II-presented epitope of interest comprises: providing
CC a nucleic acid sequence encoding the MHC Class II-presented epitope of
CC interest, the nucleic acid sequence encoding an Ii-key motif located 4-11
CC amino acids upstream from the N-terminal residue of the MHC Class II-
CC presented epitope of interest; and modifying the Ii-key motif to decrease
CC its conformance to the archetypal Ii-Key regulatory motif. Enhancing an
CC immune response directed toward an MHC Class II-presented epitope of
CC interest comprises: providing a nucleic acid sequence encoding the MHC
CC Class II-presented epitope of interest, the nucleic acid sequence lacking
CC an Ii-key motif located 4-11 amino acids upstream from the N-terminal
CC residue of the MHC Class II-presented epitope of interest; and modifying
CC the nucleic acid sequence to introduce an Ii-key motif appropriately
CC spaced from the MHC Class II-presented epitope. The protein or
CC polypeptide of interest corresponds to a protein or polypeptide encoded
CC by an infectious pathogen selected from anthrax, EBOLA, HIV or influenza,
CC preferably vaccinia virus. The non-naturally occurring protein or
CC polypeptide (I) modified by recombinant DNA techniques is useful for
CC treating infectious diseases caused or associated with infection by a
CC bacterium, virus, parasite, fungus, rickettsia or other infectious
CC agents. It is also useful for treating rheumatoid arthritis, multiple
CC sclerosis, lupus erythematosus, diabetes mellitus, myasthenia gravis,
CC autoimmune thyroiditis, scleroderma, dermatomyositis, pemphigus, asthma,
CC allergic rhinitis, topical dermatitis, colitis, cancer, psoriasis or
CC adenomas. The present sequence represents the amino acid sequence of a
CC peanut allergen Ara h 2 MHC class II-presented epitope used in the
CC invention.

```
XX SQ Sequence 156 AA;
Alignment Scores:
Pred. No.: 4.23e-87 Length: 156
Score: 819.00 Matches: 153
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 64.74% Indels: 0
DB: 8 Gaps: 0

US-10-728-051-2 (1-717) x ADO38314 (1-156)
Qy 2 CTCACCACTAGTAGCCCTCGCCCTTTCTCTCGCTGCCACGCGATCTGCGAGCGAG 61
Db 4 LeuThrileLeuValAlaLeuAlaLeuPheLeuLeuAlaAlaHisAlaSerAlaArgGln 23
Qy 62 CAGTGGGAATCCNAGGAGACAGAGATGTCAGAGCGAGTCGAGAGGGCGAACTTGAGG 121
Db 24 GlnTrpGluLeuGlnGlyAspArgArgCysGlnSerGlnLeuGluArgAlaAsnLeuArg 43
Qy 122 CCTGCGGAGCAATCTCATGCGAGAGATCCACGTCACGAGGATTCATATGACGGGAC 181
Db 44 ProCysGlnGlnHisLeuMetGlnLysileGlnArgAspGluAspSerTyrGluArgAsp 63
Qy 182 CCGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCATATATGATCGAGAGCGCTGGA 241
Db 64 ProTyrSerProSerGlnAspProTyrSerProSerProTyrAspArgArgGlyAlaGly 83
Qy 242 TCCTCTCAGCACCAAGAGAGGTGTGCAATGAGCTGAACGAGTTTGAGAACACCAAGG 301
Db 84 SerSerGlnHisGlnGluArgCysCysAsnGluLeuAsnGluPheGluAsnGlnArg 103
Qy 302 TGCATGTGCGAGGATTCGACAGATCATGAGAACGAGCGATAGCTTGCAGGGGAGG 361
Db 104 CysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArgLeuGlnGlyArg 123
Qy 362 CAACAGGAGCAACAGTTCAAGAGGGAGCTCAGGAACCTTGCTCAACAGTGGCGCTTAGG 421
Db 124 GlnGlnGlnGlnGlnPheLysArgGluLeuArgAsnLeuProGlnGlnCysGlyLeuArg 143
Qy 422 GCACCAAGCGTTGCGACTTGGAGCTGCGAAAGTGGCGGC 460
Db 144 AlaProGlnArgCysAspLeuAspValGluSerGlyGly 156

RESULT 10
ABU52576
ID ABU52576 standard; protein; 157 AA.
AC ABU52576;
XX ABU52576;
DT 10-MAR-2003 (first entry)
XX Peanut Ara h2 mutant 5 (Q20A/Q31A/D60A/D67A).
DE Allergy; Ara h1; Ara h2; Ara h3; IGE binding site; peanut; mutant;
KW mutein; anaphylactic food allergen; anti-allergenic; vaccine;
KW wound healing.
XX Homo sapiens.
OS
XX WO200274250-A2.
FN
XX 26-SEP-2002.
PD
XX 18-MAR-2002; 2002WO-US009108.
XX
XX 16-MAR-2001; 2001US-0276822P.
PR
XX 18-MAR-2002; 2002US-00276822.
XX
XX (PANA-) PANACEA PHARM.
XX
XX Caplan M, Sosin H, Sampson H, Bannon GA, Burks WA, Cockrell G;
PI
```

```
PI Compadre CM, Connaughton C, Helm RM, King NE, Kopper RA, Maleki SJ;
PI Rabjohn PA, Shin DS, Stanley JS;
XX WPI; 2003-018765/01.
DR
XX New modified anaphylactic food allergen, useful for preventing or
PT treating allergic reactions associated with e.g. anaphylactic allergens.
XX
XX Example 13; Page; 300pp; English.
PS
XX The invention relates to a modified anaphylactic food allergen has an
CC amino acid sequence that is substantially identical to that of natural
CC anaphylactic food allergen, except for a cysteine residue that has been
CC modified so that it cannot participate in the disulphide bond. The
CC modification may also comprise mutation of the IGE binding sites to
CC reduce allergenicity. Also included are: (1) a method of making a
CC modified anaphylactic food allergen; (2) a nucleotide molecule encoding
CC or for causing a site specific mutation in the modified anaphylactic food
CC allergen; (3) a transgenic plant or animal expressing the modified
CC anaphylactic food allergen; (4) a method of treating an individual by
CC reducing the clinical response to a natural anaphylactic food allergen;
CC and an isolated fragment of peanut allergen Ara h 1. The modified
CC anaphylactic food allergen is useful for preventing or treating allergic
CC reactions associated with any natural allergen such as food, insect,
CC rubber or preferably anaphylactic allergens. It is also useful for
CC treating wounds in mammals such as bovine, canine, feline, caprine,
CC ovine, porcine, murine or equine species. The present sequence is a food
CC allergen, mutated to alter its IGE binding characteristics. Note: The
CC present sequence is not shown in the specification but was created by the
CC indexer using information provided in the specification
XX
SQ Sequence 157 AA;
Alignment Scores:
Pred. No.: 1.65e-86 Length: 157
Score: 814.00 Matches: 153
Percent Similarity: 97.45% Conservative: 0
Best Local Similarity: 97.45% Mismatches: 4
Query Match: 64.35% Indels: 0
DB: 6 Gaps: 0

US-10-728-051-2 (1-717) x ABU52576 (1-157)
Qy 2 CTCACCACTAGTAGCCCTCGCCCTTTCTCTCGCTGCCACGCGATCTGCGAGCGAG 61
Db 1 LeuThrileLeuValAlaLeuAlaLeuPheLeuLeuAlaAlaHisAlaSerAlaArgAla 20
Qy 62 CAGTGGGAATCCAAAGGAGACAGAGATGCCAGAGCCAGCTCGAGAGGGCGCACTGAGG 121
Db 21 GlnTrpGluLeuGlnGlyAspArgArgCysAlaSerGlnLeuGluArgAlaAsnLeuArg 40
Qy 122 CCTGCGGAGCAATCTCATGCGAGAGATCCAACTGAGGAGGATTCATATGAACGGGAC 181
Db 41 ProCysGluGlnHisLeuMetGlnLysileGlnArgAspGluAspSerTyrGluArgAla 60
Qy 182 CCGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCATATGATCGAGAGGGCGCTGGA 241
Db 61 ProTyrSerProSerGlnAlaProTyrSerProSerProTyrAspArgArgGlyAlaGly 80
Qy 242 TCCTCTCAGCAACCAAGAGAGGTGTGCAATGAGCTGAACGAGTTTGAGAACACCAAGG 301
Db 81 SerSerGlnHisGlnGluArgCysCysAsnGluLeuAsnGluPheGluAsnGlnArg 100
Qy 302 TGCATGTGCGAGGATTCGACAGATCATGAGAACGAGCGATAGCTTGCAGGGGAGG 361
Db 101 CysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArgLeuGlnGlyArg 120
Qy 362 CAACAGGAGCAACAGTTCAAGAGGGAGCTCAGGAACCTTGCTCAACAGTGGCGCTTAGG 421
Db 121 GlnGlnGlnGlnGlnPheLysArgGluLeuArgAsnLeuProGlnGlnCysGlyLeuArg 140
Qy 422 GCACCAAGCGTTGCGACTTGGAGCTGCGAAAGTGGCGGCAGACAGATAC 472
```


141 AlaProGlnArgCysAspLeuAspValGluSerGlyArgAspArgTyr 157

RESULT 11

ID ADM12096 standard; protein; 156 AA.

XX AC ADM12096;

XX DT 20-MAY-2004 (first entry)

XX DE Arachis hypogaea 2 (Ara h2) protein.

XX KW antigen presentation enhancing hybrid polypeptide; mammalian Ii-key;
XX KW MHC class II; antibacterial; virucide; fungicide; antirheumatic;
XX KW antiarthritic; neuroprotective; dermatological; immunosuppressive;
XX KW antiinflammatory; antidiabetic; antithyroid; immune;
XX KW rheumatoid arthritis; multiple sclerosis; lupus erythematosus;
XX KW diabetes mellitus; myasthenia gravis; autoimmune thyroiditis;
XX KW scleroderma; dermatomyositis; pemphigus.

XX OS Arachis hypogaea.

XX PN US2003235594-A1.

XX PD 25-DEC-2003.

XX PF 17-SEP-2002; 2002US-00245871.

XX PR 14-SEP-1999; 99US-00396813.

XX PR 17-JUL-2002; 2002US-00197000.

XX PA (ANTI-) ANTIGEN EXPRESS INC.

XX PI Humphreys R, Xu M;

XX DR WPI; 2004-070554/07.

XX PT Novel Ii-Key/antigen presentation enhancing hybrid polypeptide, useful
XX PT for treating infections, rheumatoid arthritis, multiple sclerosis, lupus
XX PT erythematosus and diabetes mellitus.

XX PS Example 2; Page 19; 87pp; English.

XX CC The invention relates to a novel antigen presentation enhancing hybrid
XX CC polypeptide. The novel polypeptide has an N-terminal element consisting
XX CC of 4-16 residues of a mammalian Ii-Key peptide and its non-N-terminal
XX CC deletion modifications, a chemical structure covalently linking the N-
XX CC terminal element to an MHC class II-presented epitope of a C-terminal
XX CC element. The C-terminal element comprises an antigenic epitope, which
XX CC binds to an antigenic peptide binding site of an MHC class II molecule.
XX CC The antigen presentation enhancing hybrid polypeptide has the following
XX CC activities: antibacterial, virucide, fungicide, antirheumatic,
XX CC antiarthritic, neuroprotective, dermatological, immunosuppressive,
XX CC antiinflammatory, antidiabetic, and antithyroid. The antigen presentation
XX CC enhancing hybrid polypeptide is useful for modulating the immune response
XX CC in an individual and for treating infections (such as bacteria, virus,
XX CC parasite and fungus), rheumatoid arthritis, multiple sclerosis, lupus
XX CC erythematosus, diabetes mellitus, myasthenia gravis, autoimmune
XX CC thyroiditis, scleroderma, dermatomyositis and pemphigus. This sequence
XX CC represents a mammalian Ii key related protein of the invention.

XX SQ Sequence 156 AA;

Alignment Scores:

Pred. No.:	2.83e-96	Length:	156
Score:	812.00	Matches:	152
Percent Similarity:	99.35%	Conservative:	0
Best Local Similarity:	99.35%	Mismatches:	1
Query Match:	64.19%	Indels:	0
DB:	8	Gaps:	0

US-10-728-051-2 (1-717) x ADM12096 (1-156)

QY	2	CTCACCATACTAGTAGCCCTCGCCCTTTTCCTCCTCGCTCCCGCAGCATCTGCAGGACG	61
DB	4	LeuThrIleLeuValAlaLeuAlaLeuPheLeuLeuAlaAlaHisAlaSerAlaArgGln	23
QY	62	CAGTGGAACTCCCAAGAGACAGAAAGATGCGAGGCGAGCTCGAGAGGGCGAACCCTCAGG	121
DB	24	GlnTrpGluLeuGlnGlyAspArgArgCysGlnSerGlnLeuGluArgAlaAsnLeuArg	43
QY	122	CCTCGAGCAACATCTCATGCAAGATCCAAAGTCAACGTCAGAGGATTTCATATGAACGGAC	181
DB	44	ProCysGluGlyHisLeuMetGlnLysIleGlnArgAspGluAspSerTyrGluArgAsp	63
QY	182	CCGTAGACCCCTAGTCAGGATCCGTACAGCCCTAGTCCATATATCGAGAGCGCTGGA	241
DB	64	ProTyrSerProSerGlnAspProTyrSerProSerProTyrAspArgArgGlyAlaGly	83
QY	242	TCCTCTCAGCACCAAGAGAGGTGTGCAATGAGCTGAACTGAGTTCAGAAACAACCAAGG	301
DB	84	SerSerGlnHisGlnGluArgCysAsnGluLeuAsnGluPheGluAsnAsnGlnArg	103
QY	302	TGCATGTGCGAGGCATTCGAACAGATCATGGAGAACAGAGCGATAGTTGCGAGGGAGG	361
DB	104	CysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArgLeuGlnGlyArg	123
QY	362	CAACAGAGCAACAGTTCAGAGGGAGCTCAGGAAGTTCCTCAACAGTGGCGCTTAGG	421
DB	124	GlnGlnGluGlnGlnPheLysArgGluLeuArgAsnLeuProGlnGlnCysGlyLeuArg	143
QY	422	GCACCACAGCGTTGCGACTTGGACGTCGAAAGTGGCGGC	460
DB	144	AlaProGlnArgCysAspLeuAspValGluSerGlyGly	156

RESULT 12

ABU52575
ID ABU52575 standard; protein; 157 AA.

XX AC ABU52575;

XX DT 10-MAR-2003 (first entry)

XX DE Peanut Ara h2 mutant 4 (E35A/P41A/D60A/D67A).

XX KW Allergy; Ara h1; Ara h2; Ara h3; IGE binding site; peanut; mutant;
XX KW mutein; anaphylactic food allergen; antiallergenic; vaccine;
XX KW wound healing.

XX OS Homo sapiens.

XX PN WO200274250-A2.

XX PD 26-SEP-2002.

XX PF 18-MAR-2002; 2002WO-US009108.

XX PR 16-MAR-2001; 2001US-0276822P.

XX PR 18-MAR-2002; 2002US-00276822.

XX PA (PANA-) PANACEA PHARM.

XX PI Caplan M, Sosin H, Sampson H, Bannon GA, Burks WA, Cockrell G;
XX PI Compadre CM, Connaughton C, Helm RM, King NE, Kopper RA, Maleki SJ;
XX PI RabJohn PA, Shin DS, Stanley JS;

XX DR WPI; 2003-018765/01.

XX PT New modified anaphylactic food allergen, useful for preventing or
XX PT treating allergic reactions associated with e.g. anaphylactic allergens.

XX PS Example 13; Page; 300pp; English.

XX CC The invention relates to a modified anaphylactic food allergen has an
XX CC amino acid sequence that is substantially identical to that of natural
XX CC anaphylactic food allergen, except for a cysteine residue that has been

CC modified so that it cannot participate in the disulphide bond. The
CC modification may also comprise mutation of the IGE binding sites to
CC reduce allergenicity. Also included are: (1) a method of making a
CC modified anaphylactic food allergen; (2) a nucleotide molecule encoding
CC or for causing a site specific mutation in the modified anaphylactic food
CC allergen; (3) a transgenic plant or animal expressing the modified
CC anaphylactic food allergen; (4) a method of treating an individual by
CC reducing the clinical response to a natural anaphylactic food allergen;
CC and an isolated fragment of peanut allergen Ara h 1. The modified
CC anaphylactic food allergen is useful for preventing or treating allergic
CC reactions associated with any natural allergen such as food, insect,
CC rubber or preferably anaphylactic allergens. It is also useful for
CC treating wounds in mammals such as bovine, canine, feline, caprine,
CC ovine, porcine, murine or equine species. The present sequence is a food
CC allergen, mutated to alter its IGE binding characteristics. Note: The
CC present sequence is not shown in the specification but was created by the
CC indexer using information provided in the specification

XX SQ Sequence 157 AA;

Alignment Scores:
Pred. No.: 2,84e-86 Length: 157
Score: 812.00 Matches: 153
Percent Similarity: 97.45% Conservative: 0
Best Local Similarity: 97.45% Mismatches: 4
Query Match: 64.19% Indels: 0
DB: 6 Gaps: 0

US-10-728-051-2 (1-717) x ABUS2575 (1-157)

QY 2 CTCACCATAGTAGCCCTCGCCCTTTCTCTCTCGCTGCCACGCATCTCGAGGCGAG 61
DB 1 LeuThrIleLeuValAlaLeuAlaLeuPheLeuLeuAlaAlaHisAlaSerAlaArgGln 20
QY 62 CAGTGGGACTCCAGGAGACACAGATGCCAGGCGAGCTCAGAGGGCGAACCCTGAGG 121
DB 21 GlnTrpGluLeuGlnGlyAspArgCysGlnSerGlnLeuAlaArgAlaAsnLeuArg 40
QY 122 CCTGCGAGCAACATCTCATGCGAGAGATCCACGTCGACGAGATTTCATATGAACGGGAC 181
DB 41 AlaCysGluGlnHisLeuMetGlnLysIleGlnArgAspGluAspSerTyrGluArgAla 60
QY 182 CGGTACAGCCCTAGTCAGGATCCGTCAGCCCTAGTCCTATATGATCGAGAGCGCTGGA 241
DB 61 ProTyrSerProSerGlnAlaProTyrSerProSerProTyrAspArgGlyAlaGly 80
QY 242 TCCTCTCAGCACCAAGAGAGTGTTCATAGACTGAACGAGTTCAGACACCAAGG 301
DB 81 SerSerGlnHisGlnArgCysCysAsnGluLeuAsnGluPheGluAsnGlnArg 100
QY 302 TGCATGTGCGAGCATTCGCAACAGATCATGGAGAACGAGCGCATAGGTTGAGGGGAGG 361
DB 101 CysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArgLeuGlnGlyArg 120
QY 362 CAACAGAGCAACAGTTCAAGGGAGCTCAGGAATTCGCTCAACAGTCCGCGCTTAGG 421
DB 121 GlnGlnGluGlnGlnPheLysArgGluLeuArgAsnLeuProGlnGlnCysGlyLeuArg 140
QY 422 GCACCAAGCGTTGGCACTTGGACGTCGAAGTGGCGGCGAGACAGATAC 472
DB 141 AlaProGlnArgCysAspLeuAspValGluSerGlyArgAspArgTyr 157

RESULT 13
ABUS2577

ID ABUS2577 standard; protein; 157 AA.

XX AC ABUS2577;

XX 10-MAR-2003 (first entry)

DT Peanut Ara h2 mut. (W22A/E35A/P41A/D53A/D60A/D67A/R120A/L130A/L147A).
DE Allergy; Ara h1; Ara h2; Ara h3; IGE binding site; peanut; mutant;

KW Allergy; Ara h1; Ara h2; Ara h3; IGE binding site; peanut; mutant;

KW mutein; anaphylactic food allergen; antiallergenic; vaccine;
KW wound healing.
XX Homo sapiens.
XX WO200274250-A2.
XX 26-SRP-2002.
XX 18-MAR-2002; 2002WO-US009108.
XX 16-MAR-2001; 2001US-0276822P.
XX 18-MAR-2002; 2002US-00276822.
XX (PANA-) PANACEA PHARM.
XX Caplan M, Sosin H, Sampson H, Bannon GA, Burks WA, Cockrell G;
PI Compadre CM, Connaughton C, Helm RM, King NE, Kopper RA, Maleki SJ;
PI Rabjohn PA, Shin DS, Stanley JS;
XX WPI; 2003-018765/01.
XX New modified anaphylactic food allergen, useful for preventing or
PT treating allergic reactions associated with e.g. anaphylactic allergens.
XX Example 13; Page; 300pp; English.

CC The invention relates to a modified anaphylactic food allergen has an
CC amino acid sequence that is substantially identical to that of natural
CC anaphylactic food allergen, except for a cysteine residue that has been
CC modified so that it cannot participate in the disulphide bond. The
CC modification may also comprise mutation of the IGE binding sites to
CC reduce allergenicity. Also included are: (1) a method of making a
CC modified anaphylactic food allergen; (2) a nucleotide molecule encoding
CC or for causing a site specific mutation in the modified anaphylactic food
CC allergen; (3) a transgenic plant or animal expressing the modified
CC anaphylactic food allergen; (4) a method of treating an individual by
CC reducing the clinical response to a natural anaphylactic food allergen;
CC and an isolated fragment of peanut allergen Ara h 1. The modified
CC anaphylactic food allergen is useful for preventing or treating allergic
CC reactions associated with any natural allergen such as food, insect,
CC rubber or preferably anaphylactic allergens. It is also useful for
CC treating wounds in mammals such as bovine, canine, feline, caprine,
CC ovine, porcine, murine or equine species. The present sequence is a food
CC allergen, mutated to alter its IGE binding characteristics. Note: The
CC present sequence is not shown in the specification but was created by the
CC indexer using information provided in the specification

XX SQ Sequence 157 AA;

Alignment Scores:
Pred. No.: 8,63e-82 Length: 157
Score: 774.00 Matches: 148
Percent Similarity: 94.27% Conservative: 0
Best Local Similarity: 94.27% Mismatches: 9
Query Match: 61.19% Indels: 0
DB: 6 Gaps: 0

US-10-728-051-2 (1-717) x ABUS2577 (1-157)

QY 2 CTCACCATAGTAGCCCTCGCCCTTTCTCTCTCGCTGCCACGCATCTCGAGGCGAG 61
DB 1 LeuThrIleLeuValAlaLeuAlaLeuPheLeuLeuAlaAlaHisAlaSerAlaArgGln 20
QY 62 CAGTGGGACTCCAGGAGACACAGATGCCAGGCGAGCTCAGAGGGCGAACCCTGAGG 121
DB 21 GlnAlaGluLeuGlnGlyAspArgCysGlnSerGlnLeuAlaArgAlaAsnLeuArg 40
QY 122 CCTGCGAGCAACATCTCATGCGAGAGATCCACGTCGACGAGATTTCATATGAACGGGAC 181
DB 41 AlaCysGluGlnHisLeuMetGlnLysIleGlnArgAlaGluAspSerTyrGluArgAla 60
QY 182 CCGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCATATGATCGGAGAGCGCTGGA 241

```
|||||
Db 61 ProTyrSerProSerGlnAlaProTyrSerProSerProTyrAspArgArgGlyAlaGly 80
QY 242 TCCTCTCAGCACCAGAGAGGTGTGCAATCAGCTGAACGAGTTTGAGAAACCAAGG 301
Db 81 SerSerGlnHisGlnGluArgCysCysAsnGluLeuAsnGluPheGluAsnGlnArg 100
QY 302 TGATGTGCCAGGATTGCAACAGATCATCGAGAACCCAGAGCCATAGGTTGCGAGGAGG 361
Db 101 CysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArgLeuGlnGlyAla 120
QY 362 CAACAGGAGCAACAGTTCAAGAGGAGGCTCAGGAACCTTGCTCAACAGTCGGCTTAGG 421
Db 121 GlnGlnGlnGlnPheLysArgGluAlaArgAsnLeuProGlnGlnCysGlyLeuArg 140
QY 422 GCACCACAGGCTTCGCACTTGGACGTGCAAGAGTGGCGGCGAGACAGATAC 472
Db 141 AlaProGlnArgCysAspAlaAspValGluSerGlyGlyArgAspArgTyr 157

RESULT 14
ID AAY40968 standard; protein; 166 AA.
XX AC AAY40968;
XX 17-OCT-2003 (revised)
DT 06-DEC-1999 (first entry)
XX XX
XX Recombinant Ara h 2 protein sequence.
XX Peanut; allergen; Ara h 1; IgE; immunoglobulin E; epitope; Ara h 3;
KW allergic reaction; Ara h 2.
XX Arachis hypogaea.
XX WO9945961-A1.
XX 16-SEP-1999.
XX 12-MAR-1999; 99WO-US005494.
XX 12-MAR-1998; 98US-0077763P.
XX 11-MAR-1999; 99US-00077763.
XX (UYAR-) UNIV ARKANSAS.
XX Burks W, Helm RM, Cockrell G, Bannon GA, Stanley JS, Shin DS;
PI Sampson H, Compadre CM, Huang SK, Maleki SJ, Kopper RA;
XX WPI; 1999-551218/46.
XX Tertiary structure of peanut allergen Ara h 1 for protection of a host
PT animal from allergic reaction.
XX Disclosure; Page 81; 193pp; English.
XX The invention provides a tertiary structure for the peanut allergen Ara H
CC 1. The Ara H 1 allergen is found to contain 23 linear IgE-binding
CC epitopes. The invention also provides an isolated recombinant peanut
CC allergen designated Ara h 3 and a nucleotide molecule encoding the peanut
CC allergen Ara h 3. Molecules of the invention are used to protect a host
CC animal from allergic reaction, particularly using a modified allergen
CC which is less reactive with IgE. The invention may also be used to ensure
CC that the allergen is not introduced into genetically modified food. The
CC present sequence represents a recombinant Ara h 2 protein. (Updated on 17
CC -OCT-2003 to standardise OS field)
XX SQ Sequence 166 AA;

Alignment Scores:
Pred. No.: 2e-81 Length: 166
Score: 771.00 Matches: 141
Percent Similarity: 100.00% Conservative: 0
```

```
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 60.95% Indels: 0
DB: 2 Gaps: 0
US-10-728-051-2 (1-717) x AAY40968 (1-166)
QY 50 TCTGCCAGGCGAGCAGTGGGAACCTCCAAAGGAGACAGAAGATGCCAGAGCCAGCTCGAGAGG 109
Db 15 SerAlaArgGlnGlnTrpGluLeuGlnGlyAspArgArgCysGlnSerGlnLeuGluArg 34
QY 110 GCCAACTCAGGCGCTTCGCGAGCAACATCTCATCGAGAAGATCCAACTGACGAGGATTCA 169
Db 35 AlaAsnLeuArgProCysGluGlnHisLeuMetGlnLysIleGlnArgAspGluAspSer 54
QY 170 TATGAACGGGACCCGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCATATGATCGG 229
Db 55 TyrGluArgAspProTyrSerProSerGlnAspProTyrSerProSerProTyrAspArg 74
QY 230 AGAGGCGCTGGATCCCTCTCAGCACCAAGAGAGGTGTTCATGAGCTGAACGAGTTGAG 289
Db 75 ArgGlyAlaGlySerSerGlnHisGlnGluArgCysCysAsnGluLeuAsnGluPheGlu 94
QY 290 AACAAACCAAGGTGTCATGTCCGAGGCATTCGCAACAGATCATCGAGAACCAAGAGCGATAGG 349
Db 95 AsnAsnGlnArgCysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArg 114
QY 350 TTGCAGGCGGAGCAACAGGAGCAACAGTTTCAAGAGGAGCTCAGGAACTTGCCTCAACAG 409
Db 115 LeuGlnGlyArgGlnGlnGlnGlnPheLysArgGluLeuArgAsnLeuProGlnGln 134
QY 410 TCGGGCCTTAGGCGCACACAGCGTTTCGCACTTCGAGCTCGAAAGTGGCGGCGAGACAGA 469
Db 135 CysGlyLeuArgAlaProGlnArgCysAspLeuAspValGluSerGlyGlyArgAspArg 154
QY 470 TAC 472
Db 155 Tyr 155

RESULT 15
ABU52482
ID ABU52482 standard; protein; 166 AA.
XX AC ABU52482;
XX 10-MAR-2003 (first entry)
DT Peanut Ara h2 with N-terminal a T7 tag and a C-terminal His tag.
DE Peanut; allergy; Ara h1; Ara h2; Ara h3; IgE binding site;
XX anaphylactic food allergen; anti-allergenic; vaccine; wound healing.
KW Arachis hypogaea.
OS Synthetic.
OS Enterobacteria phage T7.
XX WO200274250-A2.
XX 26-SEP-2002.
XX 18-MAR-2002; 2002WO-US009108.
XX 16-MAR-2001; 2001US-0276822P.
XX 18-MAR-2002; 2002US-00276822.
XX (PANA-) PANACEA PHARM.
XX Caplan M, Sosin H, Sampson H, Bannon GA, Burks WA, Cockrell G;
PI Compadre CM, Connaughton C, Helm RM, King NE, Kopper RA, Maleki SJ;
PI Rabjohn PA, Shin DS, Stanley JS;
XX WPI; 2003-018765/01.
XX New modified anaphylactic food allergen, useful for preventing or
PT
```

PT treating allergic reactions associated with e.g. anaphylactic allergens.

XX Example 13; Fig 51; 300pp; English.

XX The invention relates to a modified anaphylactic food allergen has an amino acid sequence that is substantially identical to that of natural anaphylactic food allergen, except for a cysteine residue that has been modified so that it cannot participate in the disulphide bond. The modification may also comprise mutation of the IGE binding sites to reduce allergenicity. Also included are: (1) a method of making a modified anaphylactic food allergen; (2) a nucleotide molecule encoding or for causing a site specific mutation in the modified anaphylactic food allergen; (3) a transgenic plant or animal expressing the modified anaphylactic food allergen; (4) a method of treating an individual by reducing the clinical response to a natural anaphylactic food allergen; and an isolated fragment of peanut allergen Ara h 1. The modified anaphylactic food allergen is useful for preventing or treating allergic reactions associated with any natural allergen such as food, insect, rubber or preferably anaphylactic allergens. It is also useful for treating wounds in mammals such as bovine, canine, feline, caprine, ovine, porcine, murine or equine species. The present sequence is a peanut allergen, Ara h2, with an N-terminal T7 tag and a C-terminal His tag

SQ Sequence 166 AA;

Alignment Scores:

Pred. No.:	2e-81	Length:	166
Score:	771.00	Matches:	141
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	60.95%	Indels:	0
DB:	6	Gaps:	0

US-10-728-051-2 (1-717) x ABUS2482 (1-166)

Qy	50	TCTCGAGGCGAGCTGGAACTCCAGAGACAGAGATGCCAGAGCCAGCTCGAGAGG	109
Db	15	SerAlaArgGlnGlnTrpGluLeuGlnGlyAspArgCysGlnSerGlnLeuGluArg	34
Qy	110	CGCAACTGAGGGCCCTCGAGCAACATCTCATGAGAAGATCCAACTGACGAGGATTCA	169
Db	35	AlaAsnLeuArgProCysGluGlnHisLeuMetGlnLysIleGlnArgAspGluAspSer	54
Qy	170	TATCAACGGGACCGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCATATGATCGG	229
Db	55	TyrGluArgAspProTyrSerProSerGlnAspProTyrSerProSerProTyrAspArg	74
Qy	230	AGAGGCGCTGATCCTCTCAGCACCAAGAGAGGTGTGCAATGAGCTGACGAGTTTGAG	289
Db	75	ArgGlyAlaGlySerSerGlnHisGlnGluArgCysAsnGlnLeuAsnGluPheGlu	94
Qy	290	AACAACCAAGGTGCATGTGCGAGGCGATTGCAACAGATCATGGAGAACAGAGCCGATAGG	349
Db	95	AsnAsnGlnArgCysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArg	114
Qy	350	TTGCAGGGGAGGCAACAGGAGCAACAGTTCAAGAGGAGCTCAGGAACCTTGCTCAACAG	409
Db	115	LeuGlnGlyArgGlnGlnGlnGlnPheLysArgGluLeuArgAsnLeuProGlnGln	134
Qy	410	TGCGGCGCTTAGGGCACACAGCGCTTGCGACTTCGACGTGGAAGTGGCGGAGACACAGA	469
Db	135	CysGlyLeuArgAlaProGlnArgCysAspLeuAspValGluSerGlyArgAspArg	154
Qy	470	TAC	472
Db	155	Tyr	155

Search completed: April 8, 2005, 05:01:20
Job time : 102.896 secs

Db 24 Gln-----

```
QY 119 AGCCCTGCCAGCAACATCTCATGCAGAGATCAACGTCACGAGGATTCATATGAACGG 178
Db 40 ArgHisCysGluAsnHisIleAlaGlnArgIleGlnGlnGlnGlnGlnGluAsp 59
QY 179 GACCCGTACAGC-----CCTAGTCAGGATCCG 205
Db 60 HisAlaLeuLysLeuArgGlyIleLysHisValIleLeuArgHisArgSerGlnGlu 79
QY 206 TAGAGCCCTAGTCATATGATCGAGAGGGCGCTGATCCTCTCAGCACCAAGAGAGTGT 265
Db 80 TyrSerGlu-----GluSerGluGluLeuAspGlnCys 90
QY 266 TGCAATGAGCTGAACGAGTTTGAGAACCAACAAAGGTGTCATGTCGCGAGGATTCGAACAG 325
Db 91 CysGluGlnLeuAsnGluLeu---AsnSerGlnArgCysGlnCysArgAlaLeuGlnGln 109
QY 326 ATCATGGAGAACGAGAGCGATAGTTCAGGGAGGAGCAACAGAGCAACAGATTCAGAGG 385
Db 110 IleTyrGluSerGlnSerGlnGlnCysGluGlySerGlnGlnGlnGlnGlnGluGln 129
QY 386 GAGCTCAGGAACCTTGCTCAACAGTGCAGCTTAGGGCACCAACAGCGTTGCGACTTGGAC 445
Db 130 GluLeuGlnLysLeuProArgThrCysGlyPheGlyProLeuArgArgCysAspValAsn 149

RESULT 2
T05710
2S albumin precursor - soybean
N/alternate names: aspartic acid-rich peptide
C/Species: Glycine max (soybean)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C/Accession: T05710; A28485
R/Wang, J.; Pichersky, E.
Plant Physiol. 114, 1567, 1997
A/Title: Nucleotide Sequence of S-Adenosyl-L-Methionine:Caffeic Acid 3-O-Methyltransferase
A/Reference number: Z15424
A/Accession: T05710
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-158 <WAN>
A/Cross-references: UNIPROT:P19594; EMBL:AF005030; NID:g2305019; PID:g2305020
A/Experimental source: cultivar Hodgson 78; cotyledon
R/Odani, S.; Koide, T.; Ono, T.
J. Biol. Chem. 262, 10502-10505, 1987
A/Title: Amino acid sequence of a soybean (Glycine max) seed polypeptide having a poly(U
A/Reference number: A28485; MUID:87280104; PMID:3611081
A/Accession: A28485
A/Molecule type: protein
A/Residues: 22-64 <ODA>
C/Superfamily: soybean 2S albumin
C/Keywords: glycoprotein
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-64/Product: 2S albumin small chain #status predicted <MAT>
F;54-56/Region: cell attachment (R-G-D) motif
F;39/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:
Pred. No.: 7,86e-15 Length: 158
Score: 243.00 Matches: 56
Percent Similarity: 56.49% Conservative: 31
Best Local Similarity: 36.36% Mismatches: 55
Query Match: 19.21% Indels: 12
DB: 1 Gaps: 6

US-10-728-051-2 (1-717) x T05710 (1-158)

QY 5 ACCATAGTAGCCCTCGCCCTTTTCCTCTCGTCGCCACGCATCTGGAGGACGACG 64
Db 5 ThrIleLeuLeuIleSerLeuLeuPheCysIle---AlaHisThrCysSerAlaSerLys 23
QY 65 TGGGAACCTCCAGGAGACAGATGCGACGAGGATCGCAGAGGGCGGCACTGAGGCC 124
Db 24 TrpGlnHisGlnAsp---SerCysArgLysGlnLeuGlnGlyValAsnLeuThrPro 42
```

```
QY 125 TCGAGCAACATCTCATGCAGAGATCCAA-----CGTGACGAGGATTCATATGAACGG 178
Db 43 CysGluLysHisIleMetGluLysIleGlnGlyArgGlyAspAspAspAspAsp 62
QY 179 GACCCGTACAGCCCTAGTCAGGATCCGTACACGCCCTAGTCCATATGATCGGAGAGCGGCT 238
Db 63 AspAspAsnHisIleLeuArgThrMetArgGlyArgIleAsnTyrIleArgArgAsnGlu 82
QY 239 GGATCCTCTCAG-----CACCAGAGAGGTTGCAATGAGCTGAAC 280
Db 83 GlyLysAspGluAspGluGluGluGlyHisMetGlnLysCysCysThrGluMetSer 102
QY 281 GAGTTTGAGAACCAACCAAGGTGTCATGTCGAGGCAATTCACACAGATTCATGGAACACAG 340
Db 103 GluLeu---ArgSerProLysCysGlnCysLysAlaLeuGlnLysIleMetGluAsnGln 121
QY 341 AGCGATAGTTCAGGGAGGAGCAACAGGAGCAACAGATTCAGAGGAGGCTCAGGAACCTTG 400
Db 122 SerGluGluLeuGlu---GluLysGlnLysLysLysMetGluLysGluLeuIleAsnLeu 140
QY 401 CCTCAACAGTGCAGCCCTTAGGGCACCAACAGCGTTGCGACTTG 442
Db 141 AlaThrMetCysArgPheGlyProMetIleGlnCysAspLeu 154

RESULT 3
B23617
conglutin delta-2 large chain - narrow-leaved blue lupine
C/Species: Lupinus angustifolius (narrow-leaved blue lupine)
C/Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 09-Jul-2004
C/Accession: B23617
R/Lilley, G.G.; Inglis, A.S.
FEBS Lett. 195, 235-241, 1986
A/Title: Amino acid sequence of conglutin delta, a sulfur-rich seed protein of Lupinus ar
A/Reference number: A91358
A/Accession: B23617
A/Molecule type: protein
A/Residues: 1-80 <HIL>
A/Cross-references: UNIPROT:P09931
C/Superfamily: soybean 2S albumin

Alignment Scores:
Pred. No.: 1.42e-09 Length: 80
Score: 187.50 Matches: 34
Percent Similarity: 73.13% Conservative: 15
Best Local Similarity: 50.75% Mismatches: 17
Query Match: 14.82% Indels: 1
DB: 2 Gaps: 1

US-10-728-051-2 (1-717) x B23617 (1-80)

QY 245 TCTCAGCAACAGAGAGGTTGTCATGATGAGCTGAACGAGTTTGAGAACCAACCAAGGTGC 304
Db 11 SerGluGluLeuAspGlnCysCysGluGlnLeuAsnGluLeu---AsnSerGlnArgCys 29
QY 305 ATGTGCGAGGCAATTCACAGATCATGAGACACGAGGATGAGTTCGAGGGAGGCA 364
Db 30 GlnCysArgAlaLeuGlnGlnIleTyrGluSerGlnSerGluGlnCysGluGlyArgGln 49
QY 365 CAGGAGCAACAGTTCAAGAGGAGGCTCAGGAACTTCCTCAACAGTTCGCGCTTAGGGCA 424
Db 50 GlnGluGlnGlnLeuGluGlyGluLeuGluLysLeuProArgIleCysGlyPheGlyPro 69
QY 425 CCACAGCGTTGCGACTTGGAC 445
Db 70 LeuArgArgCysAsnIleAsn 76

RESULT 4
S38887
2S albumin - common sunflower (fragment)
C/Species: Helianthus annuus (common sunflower)
C/Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
C/Accession: S38887
```

R;Thoyts, P.J.E.; Millichip, M.; Stobart, A.K.; Griffiths, W.T.; Napier, J.A.; Shewry, J.
submitted to the EMBL Data Library, November 1993

A;Reference number: S38887
A;Accession: S38887
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-323 <THO>
A;Cross-references: UNIPROT:Q39928; EMBL:X76101; NID:g429181; PID:g429182
C;Superfamily: gliadin

Alignment Scores:
Pred. No.: 2,65e-09 Length: 323
Score: 184.50 Matches: 46
Percent Similarity: 48.99% Conservative: 27
Best Local Similarity: 30.87% Mismatches: 43
Query Match: 14.58% Indels: 33
DB: 2 Gaps: 8

US-10-728-051-2 (1-717) x S38887 (1-323)

```
QY 83 AGAAGATGC---CAGAGCCAGCTCGAGAGGCGAACCTGAGGCCCTGCGAGCAACATCTC 139
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 177 GlnGlnCysArgGlnSerGluIleGlnArgPro---ValSerGlnCysGlnArgTyrVal 195
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 140 ATCAGAGATCCAA-----CCTGAGGAG 163
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 196 GluGlnGlnIleGlnSerSerArgProTyrGlnGlnSerProTyrAspArgArgGlnGln 215
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 164 GATTCATATCAACGGGACCGTACAGCCCT-----AGTCAG 199
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 216 SerProTyrAspArgArgGlnGlnSerProTyrGluGlnArgGlnGlnProTyrGluGln 235
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 200 GATCCGTACAGCCCTAGTCCATATGATCGAGAGGCGCTGGATCTCTCTCAGCACCAA--- 256
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 236 ArgProTyrGluGlnArgProTyrGlnGlnArgGlyArgGlnGlnGlnGlnGlnGly 255
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 257 ---GAGAGGTGGTCATGAGCTGACAGAGTTTGAGAACCAACCAAGGTGCATGTGGAG 313
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 256 LeuGlnGlnCysCysAsnGluLeuGlnAsnVal-----ArgArgGluCysGlnCysGlu 273
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 314 GCATTGCAACAGATC-----ATGGAGAACAGAGCGATAGTTGCAG 355
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 274 AlaIleGlyGluValGlyGlnArgMetArgGlnGlnGlnGlnGlnGlnArgArgGlnTyr 293
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 356 GGGAGGCAACAGGAGCAACAGTTCACAGAGGAGCTCAGGAACCTTGCCTCAACAGTGGCGC 415
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 294 GlyGlyGlnGlnThrGlnThrValGluArgIleLeuGluAsnLeuProAsnGlnCysAsp 313
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 416 CTTAGGGCACCACAGCGCTGGACTTG 442
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 314 LeuAspVal---GlnGlnCysAsnIle 321
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 5

T10257
2S albumin precursor - cucurbit
N;Alternate names: prepro2S albumin
C;Species: Cucurbita sp. (cucurbit)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C;Accession: T10257; S19323
R;Hara-Nishimura, I.; Takeuchi, Y.; Inoue, K.; Nishimura, M.
Plant J. 4, 793-800, 1993
A;Title: Vesicle transport and processing of the precursor to 2S albumin in pumpkin.
A;Reference number: 217000; MUID:94100993; PMID:8275099
A;Accession: T10257
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA; protein
A;Residues: 1-141 <HAR1>
A;Cross-references: EMBL:D16560; NID:g459404; PIDN:BA003993.1; PID:g459405
A;Experimental source: seed, storage deposition stage; cotyledon
A;Note: soluble seed protein
R;Hara-Nishimura, I.; Inoue, K.; Nishimura, M.
FEBS Lett. 294, 89-93, 1991
A;Title: A unique vacuolar processing enzyme responsible for conversion of several prop

A;Reference number: S19323; MUID:92077151; PMID:1743299

A;Accession: S19323
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 26-45;65-84 <HAR2>
C;Keywords: seed; storage protein
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-35/Domain: propeptide #status experimental <PRO>
F;36-141/Product: 2S albumin #status experimental <MAT>

Alignment Scores:
Pred. No.: 2,97e-08 Length: 141
Score: 173.50 Matches: 52
Percent Similarity: 49.37% Conservative: 26
Best Local Similarity: 32.91% Mismatches: 47
Query Match: 13.72% Indels: 33
DB: 2 Gaps: 10

US-10-728-051-2 (1-717) x T10257 (1-141)

```
QY 2 CTCACCATACTAGTAGCCCTC---GCCCTTTTCTCTCTCGTCCGCCACGCATCTGCGAGG 58
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4 LeuThrSerIleIleAlaLeuPheAlaValAlaLeuValAlaAspAlaTyrAlaTyr 23
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 59 CAG-----CAGTGGGAACCTCCAAGG---GACAGAAGATGC 91
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 24 ArgThrThrIleThrThrValGluValGluGluAsnArgGlnGlyArgGluGluArgCys 43
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 92 CAGAGCCAGCTCGAGAGGCGAACCTGAGGCCCTGCGAGCAACATCTCATGACAGAAGATC 151
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 44 ArgGlnMetSerAlaArgGluLeuLeuArgSerCysGluGlnTyrLeuArgGlnGln--- 62
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 152 CAACGTGTCGAGGATTCATATGAACCGGACCCCTACAGCCCTAGT-----CAGGATCCG 205
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 63 -----SerArgAspValLeuGlnMetArgGlyIleGluAsnPro 75
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 206 TACAGCCCTAGTCCATATGATCGAGAGGCGCTGGATCTCTCAGCACCAAGAGAGGTGT 265
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 76 Trp-----ArgArgGluGlyGlySerPhe-----AspGluCys 86
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 266 TGCATGAGTGAACAGAGTTTGAGAACCAACAAAGGTGCATGTGCGAGGATTCGCAACAG 325
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 87 CysArgGluLeuLeuAsnValAsp-----GluGluCysArgCysAspMetLeuGluGlu 104
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 326 ATCATCGAGAACACAGACCGCATAGTTGTCAGGGGAGGCAACAGGACCAACAGTTCAGAGG 385
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 105 IleAlaArgGluGluGlnArgGlnAlaArgGly---GlnGluGlyArgGlnMetLeuGln 123
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 386 GAGCTCAGGAACCTTGCTCAACAGTCCGGCTTAGGGCACCACAGCGTTGCGAC 439
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 124 LysAlaArgAsnLeuProSerMetCysGlyIleArg---ProGlnArgCysAsp 140
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 6

S14947
2S albumin - Brazil nut
C;Species: Bertholletia excelsa (Brazil nut)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000
C;Accession: S14947
R;Gander, E.S.; Holmstroem, K.O.; de Paiva, G.R.; de Castro, L.A.B.; Carneiro, M.; Gross
Plant Mol. Biol. 16, 437-448, 1991
A;Title: Isolation, characterization and expression of a gene coding for a 2S albumin fr
A;Reference number: S14946; MUID:91370890; PMID:1840683
A;Accession: S14947
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-154 <CAN>
A;Cross-references: EMBL:X54491; NID:g17712; PIDN:CAA38363.1; PID:g17713
C;Genetics:
A;Introns: 67/3
C;Superfamily: wheat alpha-amylase inhibitor
Alignment Scores:
Pred. No.: 2,07e-06 Length: 154

Score: 154.00 Matches: 45
Percent Similarity: 45.34% Conservative: 28
Best Local Similarity: 27.95% Mismatches: 52
Query Match: 12.17% Indels: 36
DB: 2 Gaps: 10

US-10-728-051-2 (1-717) x S14947 (1-154)

```
QY 8 ATACTAGTAGCCCTCGCCCTTTCTCTCTGCTGCCACGCGATCTGCG----- 55
D 7 ValAlaAlaLeuLeuAlaLeuValLeuGluValLeuGluAlaPheArgThrThr 26
QY 56 -----AGCAGCAGTGGGAATCCCAAGCA-----GACAGAAGA 88
D 27 ValThrThrThrLeuGluGluGluGluGluGluGluGluGluGluGluGlu 46
QY 89 TGCCAGAGCCAGCTCGAGAGG---GCGAAGCTGAGGCCCTCGAGCAACATCTCATGCG 145
D 47 CysArgGluGluMetGluArgGluGluGluGluGluGluGluGluGluGlu 66
QY 146 AAGATCAACAGTGCAGAGGATTCATATCAACGGGACCCGTACAGCCCTAGTCAGGATCCG 205
D 67 GlnMet-----GluGluSerProTyrGln----- 74
QY 206 TACAGCCCTAGTCCATATGATCGAGAGCGCTGATCTCTCAGCAACCAAGAGAGGTGT 265
D 75 ---AsnProArgProLeuArgArg-----GlyGluGluProHisLeuAspGluCys 90
QY 266 TGCAATCAGCTGACGAGTGTGAGAACCAACAAAGGTGCGATGCGAGGATTCGCAACAG 325
D 91 CysGluGluGluGluArgMetAsp-----GluMetCysArgCysGluGluLeuArgMet 108
QY 326 ATCATGAGAACCAAGCGCAT-----AGGTTCAGGGAGGCAACAGAGCAACAGTTC 379
D 109 MetLeuArgArgGlnArgGluGluMetGluLeuGlnGly---GluGlnMetGlnArgGlie 127
QY 380 AAGAGGAGCTCAGGAACCTTCCTCAACAGTGGCGCTTAGGGCACCACAGCGTTGCGAC 439
D 128 MetArgLysAlaGluAsnLeuLeuSerArgCysAsnLeu---SerProGlnArgCysPro 146
QY 440 TTG 442
D 147 Met 147
```

RESULT 7

S01062
2S seed storage protein precursor (clone HaG5) - common sunflower
N:Alternate names: 2S albumin storage protein
C:Species: Helianthus annuus (common sunflower)
C>Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 09-Jul-2004
C:Accession: S01062
R:Allen, R.D.; Cohen, E.A.; Vonder Haar, R.A.; Adams, C.A.; Ma, D.P.; Nessler, C.L.; Tho
Mol. Gen. Genet. 210, 211-218, 1987
A:Title: Sequence and expression of a gene encoding an albumin storage protein in sunfl
A:Reference number: S01062; MUID:88142538; PMID:2830455
A:Accession: S01062
A:Molecule type: DNA
A:Residues: 1-295 <ALL>
A:Cross-references: UNIPROT:P15461; EMBL:X06410; NID:g18799; PIDN:CAA29699.1; PID:g18800
A>Note: part of this sequence, including the amino end of the mature protein, was confir
C:Genetics:
A:Introns: 192/2
C:Superfamily: Gliadin
F:1-20/Domain: signal sequence #status predicted <SIG>
F:162-295/Product: 2S seed storage protein #status experimental <NAT>

Alignment Scores:

Pred. No.:	3 93e-06	Length:	295
Score:	151.00	Matches:	47
Percent Similarity:	40.39%	Conservative:	35
Best Local Similarity:	23.15%	Mismatches:	65
Query Match:	11.94%	Indels:	56
DB:	2	Gaps:	8

US-10-728-051-2 (1-717) x S01062 (1-295)

```
QY 8 ATACTAGTAGCCCTCGCCCTTTCTCTCTGCTGCCACGCGATCTGCG----- 55
D 5 IleValLeuAlaLeuAlaPheAlaLeuValAlaPheAlaThrAlaHisThrThrIle 24
QY 56 -----AGCAGCAGTGGGAATCCCAAGGAGACAGAGATGCCAGGCCAG 100
D 25 IleThrThrThrIleGluAspGluAsnProIleSerGlyGlnArgGlnValSerGlnArg 44
QY 101 CTCGAGAGCGCGAACCCTCGAGCGCTCGAGCAACATCTCTCATG-----CAGAAGATCCAA 154
D 45 IleGlnGlyGlnArgLeuAsnGlnCysArgMetPheLeuGlnGlnGlyGlnAsnIlePro 64
QY 155 CGTGACGACGATTCATATGAACGGGACCCGTACAGCCCTAGTCAGATCCGTACAGCCCT 214
D 65 ArgGluPheAspAsn----- 69
QY 215 AGTCCATATGATCGGAGAGCGCTGATCTCTCAGCAACCAAGAGAGGTGTTGCAATGAG 274
D 70 -----ProGlnMetGlyArgGlnGlnGluGlnGlnLeuGlnGlnCysCysGlnGlu 86
QY 275 CTCGAACGAGTTTGAGAACCAACCAAGGTGTCATGTGCGAGGCATTCGCAACAGATCGGAG 334
D 87 LeuGlnAsnIleGluGly-----GlnCysGlnCysGluAlaValLysGlnValPheArg 104
QY 335 AACCAAGCGATGAGTTGCGAG-----GGGAGG 361
D 105 GluAlaGlnGlnGlnValGlnGlnGlnGlnGlnGlnLeuValProPheArgGlySer 124
QY 362 CAACAGGAGCAACAGTTCAAGAGGAGGAGCTTCAGGAACCTTCCTCAACAGTGGCGCTTAGG 421
D 125 GlnGlnThrGlnGlnLeuLysGlnLysAlaGlnIleLeuProAsnValCysAsnLeuGln 144
QY 422 GCACCACACGCTTGGCAGTTG-----GACGTCGAA 451
D 145 Ser---ArgArgCysGluIleGlyThrIleThrThrThrValThrGluSerAsnIleAsp 163
QY 452 AGTGGCGGACAGACAGATACCTAAACACTATCTCAAAAAAGAAAAAGAAAAAGAA 511
D 164 IleProPheArgAspArg-----PropheGlyThrGlySerGlnGlnCysArgGlu 180
QY 512 AATAGCTTA 520
D 181 ThrGluIle 183
```

RESULT 8

A29802
napin precursor (gNa) - rape
N:Alternate names: 1.7 S seed storage protein
C:Species: Brassica napus (rape)
C>Date: 20-Jun-1989 #sequence_revision 20-Jun-1989 #text_change 09-Jul-2004
C:Accession: A29802
R:Scotfield, S.R.; Crouch, M.L.
J. Biol. Chem. 262, 12202-12208, 1987
A:Title: Nucleotide sequence of a member of the napin storage protein family from Brassic
A:Reference number: A29802; MUID:87308225; PMID:3040733
A:Accession: A29802
A:Molecule type: DNA
A:Residues: 1-186 <SCO>
A:Cross-references: UNIPROT:P09893; GB:J02782; NID:g167178; PIDN:AAA33007.1; PID:g167179
C:Genetics:
A:Gene: gNa
C:Superfamily: wheat alpha-amylase inhibitor

Alignment Scores:

Pred. No.:	9 49e-06	Length:	186
Score:	147.00	Matches:	47
Percent Similarity:	40.98%	Conservative:	28
Best Local Similarity:	25.68%	Mismatches:	52
Query Match:	11.62%	Indels:	56
DB:	2	Gaps:	9

Db 70 Pro-----ArgArg-----GlyMetGluProHisMetSerGluCysCysGluGlnLeu 85
 QY 278 AACGAGTTTGAGAAACAACCAAGGTGCATGTCGAGGCAATTGCACAA--CAGATCATGGAG 333
 Db 86 GluGlyMetAsp-----GluSerCysArgCysGluGlyLeuArgMetMetMetMetArg 103
 QY 335 AACGAGGAGCATAGTTGTCGGGGAGG---CAACAGGAGCAACAGTTTCAAGAGGGAGCTC 391
 Db 104 MetGlnGlnGluGluMetGlnProArgGlyGluGlnMetArgAGMetMetArgLeuAla 123
 QY 392 AGGAACCTTGCTCAACAGTGGCGCTTAGGCGCACACAGCGTTCGCGACTTG 442
 Db 124 GluAsnIleProSerArgCysAsnLeu---SerProMetArgCysProMet 139

RESULT 11
 JC5379
 mabinlin II precursor - Yunnan caper
 C:Species: Capparis masaiakai (Yunnan caper)
 C:Date: 02-Jun-1997 #sequence_revision 12-Sep-1997 #text_change 09-Jul-2004
 A:Accession: JC5379; PC4316
 R:Nirasawa, S.; Masuda, Y.; Nakaya, K.; Kurihara, Y.
 Gene 181, 225-227, 1996
 A:Title: Cloning and sequencing of a cDNA encoding a heat-stable sweet protein
 A:Reference number: JC5379; MUID:97128796; PMID:8973336
 A:Accession: JC5379
 A:Molecule type: mRNA
 A:Residues: 1-155 <NIRI>
 A:Cross-references: UNIPROT:P30233; DDBJ:D83997; NID:g1817545; PIDN:BAAL2204.1
 A:Accession: PC4316
 A:Molecule type: protein
 A:Residues: 36-41;149-154 <NIR2>
 A:Experimental source: seed
 C:Superfamily: wheat alpha-amylase inhibitor
 C:Keywords: sweet taste
 F;1-20/Domain: signal sequence #status predicted <SIG>
 F;21-35/Domain: amino-terminal propeptide #status predicted <PRO>
 F;36-68,83-154/Product: mabinlin II #status experimental <MAT>
 F;36-68/Domain: mabinlin II A chain #status experimental <CHA>
 F;69-83/Domain: mabinlin linker peptide #status predicted <LNK>
 F;83-154/Domain: mabinlin II B chain #status experimental <CHB>

Alignment Scores:
 Pred. No.: 5,44e-05 Length: 155
 Score: 139.00 Matches: 45
 Percent Similarity: 45.51% Conservative: 26
 Best Local Similarity: 28.85% Mismatches: 47
 Query Match: 10.99% Indels: 38
 DB: 2 Gaps: 9

US-10-728-051-2 (1-717) x JC5379 (1-155)
 QY 8 ATACTAGTAGCCCTCGCCCTTCTCTCGCTGCCACGCATCTCGAGG----- 58
 Db 7 LeuPheAlaThrLeuAlaLeuPheValLeuAlaAsnAlaSerIleGlnThrThrVal 26
 QY 59 -----CAGAGTGGGAATCCACGAGGACAGAGAAGATCC 91
 Db 27 IleGluValAspGluGluAspAsnGlnLeuTrp-----ArgCys 40
 QY 92 CAGAGCCAG---CTCGAGAGGGCGCAACTCGAGGCCCTCGGAGCAACATCTCATGCGAGAG 148
 Db 41 GlnArgGlnPheLeuGlnHisGlnArgLeuArgAlaCysGlnArgPheIleHisArg 60
 QY 149 ATCCAA-----CGTGACGAGGATTTCATGAACGGGACCCGTCAGCCCTAGTCAGGAT 202
 Db 61 AlaGlnPheGlyGlyGlnProAspGluLeuGluAspGluValGluAspAsnAsp 80
 QY 203 CGGTACAGCCCTAGTCCATATGATCGAGAGGGCGCTGGATCTCTCTCAGCACCAAGAGG 262
 Db 81 GluAsnGlnPro-----ArgArgProIla-----LeuArgGln 91
 QY 263 TGTTCGATGAGCTGAACGAGTTGTGAGAACACCAAAAGTGATGTGC----- 310

Db 92 CysCysAsnGlnLeuArgGlnValAsp-----ArgProCysValCysProValLeuArg 109
QY 311 GAGCCATTGCAACAGATCATGAGAACAGAGGAGATAGTTGAGGGGAGGCAACAGAG 370
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 110 GlnAlaAlaGlnGlnValLeuArgGlnLe-----lleGlnGlyProGlnGlnLeu 127
QY 371 CAACAGTTCAAGAGGAGCTCAGAACTTCCTCAACAGTGGCGGCTT 418
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 128 ArgArgLeuPheAspAlaAlaArgAsnLeuProAsnIleCysAsnIle 143
RZCS
2S seed storage protein precursor - castor bean
N;Alternate names: 2S albumin precursor
C;Species: Ricinus communis (castor bean)
C;Date: 14-Nov-1983 #sequence revision 08-Feb-1996 #text change 09-Jul-2004
C;Accession: S11499; S11500; S11501; S11502; S27221; A01328; S27222
R;Irwin, S.D.; Lord, J.M., 1990
Nucleic Acids Res. 18, 5890, 1990
A;Title: Nucleotide sequence of a Ricinus communis 2S albumin precursor gene.
A;Reference number: S11499; MUID:91016940; PMID:2216785
A;Accession: S11499
A;Molecule type: DNA
A;Residues: 1-258 <IRW>
A;Cross-references: UNIPROT:P01089; EMBL:X54158; NID:921067; PID:CAA38097.1; PID:921068
A;Note: the authors translated the codon CTC for residue 14 as Phe, CCA for residue 74 as
R;Irwin, S.D.; Keen, J.N.; Findlay, J.B.C.; Lord, J.M.
Mol. Gen. Genet. 222, 400-408, 1990
A;Title: The Ricinus communis 2S albumin precursor: a single preproprotein may be processed
A;Reference number: S11500; MUID:91109729; PMID:2274038
A;Accession: S11500
A;Molecule type: mRNA
A;Residues: 1-13,'F',15-73,'T',75-258 <IR2>
A;Experimental source: clone 14g4
A;Accession: S11501
A;Molecule type: mRNA
A;Residues: 'N',4,'LS',7-13,'F',15-21 <IRF>
A;Experimental source: clone 10a12
A;Accession: S27221
A;Molecule type: protein
A;Residues: 'X',37,'X',39-45,'X',158-161,'X',163-174,'X', <IRA>
R;Sharief, F.S.; Li, S.S.L.
J. Biol. Chem. 257, 14753-14759, 1982
A;Title: Amino acid sequence of small and large subunits of seed storage protein from Ricinus communis
A;Reference number: A92357; MUID:83082772; PMID:7174664
A;Accession: A01328
A;Molecule type: protein
A;Residues: 157-190:194-221,'Q',223-225,230-233,'N',235-254,'Q',256-258 <SHA>
A;Note: 230-Ser was also found
A;Note: there is considerable similarity between residues 181-231 of this protein and residues 181-231 of the protein from Ricinus communis
R;Odani, S.; Koide, T.; Ono, T.; Ohnishi, K.
Biochem. J. 213, 543-545, 1983
A;Title: Structural relationship between barley (Hordeum vulgare) trypsin inhibitor and soybean trypsin inhibitor
A;Reference number: A90322; MUID:83308577; PMID:6615448
A;Contents: annotation
A;Note: this protein is homologous with trypsin inhibitor from barley
C;Complex: consists of two chains linked by two disulfide bonds involving Cys-162 and Cys-163
C;Superfamily: 2S seed storage protein precursor
C;Keywords: pyroglytamic acid; seed; storage protein
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-35/Domain: propeptide #status predicted <PRO>
F;36-72/Product: probable 2S seed storage protein small chain 2 #status experimental <SC>
F;157-190/Product: probable 2S seed storage protein large 2 #status predicted <LCHI>
F;157-190/Product: 2S seed storage protein small chain #status experimental <SML>
F;194-258/Product: 2S seed storage protein large chain #status experimental <LRG>
F;194/Modified site: pyroglutamic acid (Gln) (in mature form) #status experime
Alignment Scores:
Pred. No.: 5.38e-05 Length: 258

Score: 139.00 Matches: 40
Percent Similarity: 48.03% Conservative: 33
Best Local Similarity: 26.32% Mismatches: 51
Query Match: 10.99% Indels: 28
DB: 1 Gaps: 7
US-10-728-051-2 (1-717) x RZCS (1-258)
QY 11 CTAGTAGCCCTCGCCCTTTTCTCTCTCGTCCCGCCGATCTCGAGGCGAGCAGTGG--- 67
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 10 LeuValSerValLeuLeuPheIleIleAlaAsnAlaSerPheAlaTyrArgThrThrIle 29
QY 68 -----GAATCCAGAGAGACAGA-----AGATGC 91
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 30 ThrThrIleGluIleAspGluSerLysGlyGluArgGluGlySerSerSerGlnGlnCys 49
QY 92 CAGAGCCAGCTCGAGAGGCGACCTGAGCCCTCGAGCAACATCTCATGACAGAGATC 151
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 50 ArgGlnGluValGlnArgLysAspLeuSerSerCysGluArgTyrLeuArgGlnSerSer 69
QY 152 CAACGTGACGAGGATTTCATATGAACGGGACCCGCTACAGCCCTAGTCAGGATCCGTACAGC 211
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 70 SerArg-----ArgSerProGlyGluGluValLeuArg 80
QY 212 CCTAGTCCATATGATCGGAGAGCGGTGGATCTCTCAGCACCAAGAGAGGTGTGCAAT 271
Db 81 ---MetProGlyAspGluAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 99
QY 272 GAGCTCAACGAGTTTCAGAGAACAAAGTGTGATGCGAGGCGCATTTGCAACAGATCATG 331
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 100 GlnValLysGlnValArgAsp-----GluCysGlnCysGlnAlaIleIleTyrIleAla 117
QY 332 GAGAAC-----CAGAGCGATATGTCAGGAGGAGGACCAAGAGAGAGGTTTCAAGAGG 385
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 118 GluAspGlnIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 137
QY 386 GAGCTCAGAACTTGGCTCAACAGTGGCGGCTTAGG 421
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 138 ---AlaGlyGluIleValSerSerCysGlyValArg 148
RESULT 13
T09252
seed storage protein EMB25 - white spruce
C;Species: Picea glauca (white spruce)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C;Accession: T09252
R;Dong, J.Z.; Dunstan, D.I.
submitted to the EMBL Data Library, June 1996
A;Description: Gene expression during somatic embryogenesis.
A;Reference number: Z16588
A;Accession: T09252
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-165 <DON>
A;Cross-references: UNIPROT:Q40850; EMBL:L47745; NID:gl350513; PID:gl350514
C;Genetics:
A;Gene: EMB25
Alignment Scores:
Pred. No.: 0.000224 Length: 165
Score: 132.50 Matches: 47
Percent Similarity: 38.42% Conservative: 21
Best Local Similarity: 26.55% Mismatches: 44
Query Match: 10.47% Indels: 65
DB: 2 Gaps: 10
US-10-728-051-2 (1-717) x T09252 (1-165)
QY 8 ATACTAGTACCCCTCGCCCTTTTCTCTCTCGCT-----GCC 43
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 17 ValSerLeuGlyMetAlaLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 36
QY 44 CACGCATCTCGAGGCGAGCAGTGGGAACTCCAAGAGAGACAGAAG----TGCCAGAGCCAG 100

```

Db      37 HisGluAspGlyLeuTyrGlyGluValGlnGlnArgArgSerCysGluGlnGln 56
QY      101 CTCGAGAGGCGCAACTGAGGCCCTCGGAGCAACATCTCATGCAGAAATCCACGTGAC 160
Db      57 -----ArgLeuSerSerCysArgGluTyrLeu-----GluArgPro 68
QY      161 GAGGATTTCATATGAACGGGACCCGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCCA 220
Db      69 ArgAsp-----
QY      221 TATGATCGAGAGGCGCTGGATCTCTCAGCACCAAGAGAGGTGTTCATAGCTGAAC 280
Db      71 -----GlnProSerGluArgCysCysGluGluGln 81
QY      281 GAGTTTGAGAACCAACAAAGGTGCATGTCGAGGCATTCGCAACAGATCATGGAAAC--- 337
Db      82 ArgMet-----SerProGlnCysArgCysGlnAlaIleGlnArgThrLeuGluAspVal 99
QY      338 -----CAGAGCGAT-----AGG 349
Db      100 PheMetAspSerAspSerGlnAspGlyAlaProLeuAsnGlnArgArgGlnArg 119
QY      350 TTGAGGGGAGG---CAACAGAGCAACAGTTCAAGAGGGAGCTCAGAACTTGCCTCAA 406
Db      120 GlyGlnGlyArgGlyMetGluGluGluValValArgArgAlaGluGluLeuProAsn 139
QY      407 CAGTGGCGCCCTAGG---GCACCAACAGCTTGGCCATTCGAGCGTCGAAAGT 454
Db      140 ThrCysAsnValArgGlnSerProArgArgCysAspLeuGlnArgHisSer 156

RESULT 14
T08012
2S seed storage protein precursor - Douglas fir
C:Species: Pseudotsuga menziesii (Douglas fir)
C>Date: 21-May-1999 #sequence_revision 21-May-1999 #text_change 09-Jul-2004
C:Accession: T08012
R:Chatthai, M.; Misra, S.
Planca 206, 138-145, 1998
A:Title: Structure and expression of embryogenesis-specific cDNAs encoding 2S seed storage
A:Reference number: 216286; MUID:98381325; PMID:9715536
A:Accession: T08012
A:Status: preliminary; translated from GR/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-170 <CH>
A:Cross-references: UNIPROT:O64931; EMBL:AF0299372; NID:g2988481; PIDN:AAC27000.1; PID:g2
C:Genetics:
A:Gene: 2S3
C:Keywords: seed; storage protein

```

```

Alignment Scores:
Pred. No.: 0.00048 Length: 170
Score: 129.00 Matches: 43
Percent Similarity: 36.11% Conservative: 22
Best Local Similarity: 23.89% Mismatches: 49
Query Match: 10.20% Indels: 66
DB: 2 Gaps: 8

US-10-728-051-2 (1-717) x T08012 (1-170)
QY      8 ATACTAGTACGCTCGCCCTTTCTCTCGTGTGCCACGCTATCGGAGGAGCAGTGG 67
Db      17 ValSerLeuGlyValAlaLeuLeuLeuVal-----GlnTrp 29
QY      68 -----GAACTCAAAGGAGACAGAAAGATGCCAGAGC 97
Db      30 SerThrProAsnValAspAlaGlyAspAsnMetPheGlyGluAspValValGlnGln 49
QY      98 CAGCTCGAGAGGCGCAACCTGAGGCCCTCGGAGCAACATCTCATGCAGAGATCCAACT 157
Db      50 GlnGlnArgArgGlySerCysAspPro-----GlnArgLeuSerSer 63
QY      158 GACGAGGATTTCATGAACGGGACCCGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGT 217

```

```

Db      64 CysArgAspTyrLeuGluArgArgArgGluGlnProSer----- 76
QY      218 CCATATGATCGGAGAGCGCTGGATCCTCTCAGCACCAAGAGAGGTGTTCGAATGAGCTG 277
Db      77 -----GluArgCysCysAsnGluLeu 83
QY      278 AACGAGTTTGAGAACCAACAAAGGTGCATGTCGAGGCATTCGCAACAGATCATGCAGAAC 337
Db      84 GluArgMet-----SerProGlnCysArgCysProAlaIleGlnGlnValPheAspGln 101
QY      338 CAGAGCGATAGGTG----- 352
Db      102 SerSerGluAspLeuSerMetValAspSerHisSerGlnAsnAlaIleGlnArg 121
QY      353 -----CAGGGAGGCAACAG---GAGCAACAGTTCAAGAGGAGCTCAGGAAC 397
Db      122 ArgArgGluGluArgGlyArgGluAlaGluGluMetValGluArgAlaGlnArg 141
QY      398 TTGCCTCAACAGTGGCGCTTAGGGCACCAAGCGT---TGGACTTGGACGTGAAAAGT 454
Db      142 LeuProAsnThrCysAsnValArgGlnProProArgHisCysAspIleGlnArgHisSer 161

RESULT 15
S18871
2S-like storage protein - white spruce
C:Species: Picea glauca (white spruce)
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Jul-2004
C:Accession: S18871
R:Newton, C.H.
submitted to the EMBL Data Library, November 1991
A:Reference number: S18871
A:Accession: S18871
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-172 <NEW>
A:Cross-references: UNIPROT:P26986; EMBL:X63193; NID:g20495; PID:g20496

Alignment Scores:
Pred. No.: 0.00056 Length: 172
Score: 128.00 Matches: 46
Percent Similarity: 37.43% Conservative: 21
Best Local Similarity: 25.70% Mismatches: 46
Query Match: 10.12% Indels: 66
DB: 2 Gaps: 9

US-10-728-051-2 (1-717) x S18871 (1-172)
QY      14 GTAGCCCTCGCCCTTTCTCTCTCTC-----GTCGCCACGCA 49
Db      18 LeuSerValAlaLeuPheLeuLeuPheHisTrpGlyLeProSerValAspGlyHisGlu 37
QY      50 TCTGCGAGGAGCAGTGGGAACCTCAAGGAGACAGAGA---TGCCAGAGCCAGCTCGAG 106
Db      38 AspAsnMetTyrGlyGluGluIleGlnGlnArgSerCysAspProGlnArgAsp 57
QY      107 AGGCGCAACCTGAGCGCCCTCGAGCAACATCTCATGTGAGAGATCCAACTGACGAGAT 166
Db      58 ProGlnArgLeuSerSerCysArgAspTyrLeu----- 68
QY      167 TCATATGAACGGGACCCGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCCATATGAT 226
Db      69 -----GluArgArgGluGlnProSer----- 76
QY      227 CGGAGAGCGCTGGATCCTCTCAGCACCAAGAGAGGTGTTCGAATGAGTCAACAGAGTTT 286
Db      77 -----GluArgCysCysGluGluLeuGlnArgMet 86
QY      287 GAGAACCAACAAAGGTGCATGTCGAGGCATTCGCAACAGATCATGGAG----- 334
Db      87 -----SerProGlnCysArgCysGlnAlaIleGlnGlnMetLeuAspGlnSerLeuSer 104
QY      335 -----AACGAGCGGATAGG 349

```

```
Db      105 TyrAspSerPheMetAspSerAspSerGlnGluAspAlaProLeuAsnGlnArgArgArg 124
QY      350 TTG-----CAGGGAGGCAACAGGAGCAACAGTTCAAGAGGGAGCTCAGGAACTTG 400
Db      125 ArgArgGluGlyArgGlyArgGluGluGluAlaMetGluArg---AlaAlaTyrLeu 143
QY      401 CCTCAACAGTGGCGCCTTAGG---GCACCACAGCGTTGCGACTTGGACGTCGAAAGT 454
Db      144 ProAsnThrCysAsnValArgGluProProArgArgCysAspIleGlnArgHisSer 162
```

Search completed: April 8, 2005, 05:30:20
Job time : 36.3307 secs

This Page Blank (uspto)

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

```

QY 122 CCTGCGAGCAATCTCATGCGAGAGATCCAAACGTCGACGAGATTTCATATGAACGGGAC 181
Db 63 ProCysGluGlnHisLeuMetGlnLysIleGlnArgAspGluAspSerTyrGluArgAsp 82
QY 182 CCGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCATATCATCGGAGGCGCTGGA 241
Db 83 ProTyrSerProSerGlnAspProTyrSerProSerProTyrAspArgGlyAlaGly 102
QY 242 TCCTCTCAGCACCAAGAGAGCGTGTTCATAGCTGAACGAGTTTCAGAACCAACCAAGG 301
Db 103 SerSerGlnHisGlnGluArgCysCysAsnGluLeuAsnGluPheGluAsnGlnArg 122
QY 302 TGCATGTCCAGGCGATTGCAACAGATCATCGAGAAACAGAGCGATAGTTTCAGGGGAGG 361
Db 123 CysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArgLeuGlnGlyArg 142
QY 362 CACAGGAGCAACAGTTCAAGAGGGAGCTCAGAACTTCCTCAACAGTCGGCCCTAGG 421
Db 143 GlnGlnGlnGlnGlnPheLysArgGluLeuArgAsnLeuProGlnGlnCysGlyLeuArg 162
QY 422 GCACCCAGCGTTGCCACTTGGACGTCGAAAGTGGCGGCAGACAGATAC 472
Db 163 AlaProGlnArgCysAspLeuAspValGluSerGlyArgAspArgTyr 179

RESULT 2
Q6PSU1
ID Q6PSU1 PRELIMINARY; PRT; 175 AA.
AC Q6PSU1;
DT 05-JUL-2004 (T-EMBLrel. 27, Created)
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
DE Seed storage protein SSP2 (Fragment).
OS Arachis hypogaea (Peanut).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;
OC Arachis.
OX NCBI_TaxID=3818;
RN [1]
RP SEQUENCE FROM N.A.
RA Yan Y., Wang L., Liao B., Li H., Lin X., Huang S.;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY581854; AAT00599.1; -.
DR InterPro; IPR003612; AAI.
DR Pfam; PF00234; Tryp_alpha_amyl; 1.
DR SMART; SM00499; AAI; 1.
FT NON TER 1
SQ SEQUENCE 175 AA; 20463 MW; B3FB5806B70733DE CRC64;

Alignment Scores:
Pred. No.: 5.77e-65 Length: 175
Score: 838.00 Matches: 156
Percent Similarity: 99.36% Conservative: 0
Best Local Similarity: 99.36% Mismatches: 1
Query Match: 66.25% Indels: 0
DB: 2 Gaps: 0

US-10-728-051-2 (1-171) x Q6PSU1 (1-175)

QY 2 CTCACCATAGTAGCCCTCGCCCTTTCTCTCGTCCCGCCAGCATCTCGAGGCGAG 61
Db 19 LeuThrIleLeuValAlaLeuAlaLeuPheLeuLeuAlaHisAlaSerAlaArgGln 38
QY 62 CAGTGGGAACCTCCAAAGAGACAGAATGCCAGGCGAGCTCCAGAGGGCGAACCTGAGG 121
Db 39 GlnTrpGluPheGlnGlyAspArgArgCysGlnSerGlnLeuGluArgAlaAsnLeuArg 58
QY 122 CCTGCGAGCAATCTCATGCGAGAGATCCAAACGTCGACGAGATTTCATATGAACGGGAC 181
Db 59 ProCysGluGlnHisLeuMetGlnLysIleGlnArgAspGluAspSerTyrGluArgAsp 78
QY 182 CCGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCATATCATCGGAGGCGCTGGA 241

```

```

Db 79 ProTyrSerProSerGlnAspProTyrSerProSerProTyrAspArgGlyAlaGly 98
QY 242 TCCTCTCAGCACCAAGAGAGCGTGTTCATAGCTGAACGAGTTTCAGAACCAACCAAGG 301
Db 99 SerSerGlnHisGlnGluArgCysCysAsnGluLeuAsnGluPheGluAsnGlnArg 118
QY 302 TGCATGTCCAGGCGATTGCAACAGATCATCGAGAAACAGAGCGATAGTTTCAGGGGAGG 361
Db 119 CysMetCysGluAlaLeuGlnGlnIleMetGluAsnGlnSerAspArgLeuGlnGlyArg 138
QY 362 CACAGGAGCAACAGTTCAAGAGGGAGCTCAGAACTTCCTCAACAGTCGGCCCTAGG 421
Db 139 GlnGlnGlnGlnGlnPheLysArgGluLeuArgAsnLeuProGlnGlnCysGlyLeuArg 158
QY 422 GCACCCAGCGTTGGACGTCGAAAGTGGCGGCAGACAGATAC 472
Db 159 AlaProGlnArgCysAspLeuAspValGluSerGlyArgAspArgTyr 175

RESULT 3
Q941R0
ID Q941R0 PRELIMINARY; PRT; 156 AA.
AC Q941R0;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DE Allergen II (Fragment).
OS Arachis hypogaea (Peanut).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;
OC Arachis.
OX NCBI_TaxID=3818;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-21192158; PubMed=11295663; DOI=10.1067/mai.2001.113522;
RA Viquez O.M., Summer C.G., Dodo H.W.;
RT "Isolation and molecular characterization of the first genomic clone
of a major peanut allergen, Ara h 2.";
RL J. Allergy Clin. Immunol. 107:713-717(2001).
DR EMBL; AY007229; AAK96887.1; -.
DR InterPro; IPR003612; AAI.
DR Pfam; PF00234; Tryp_alpha_amyl; 1.
DR SMART; SM00499; AAI; 1.
FT NON TER 156
SQ SEQUENCE 156 AA; 18167 MW; 664D6832F91F36D1 CRC64;

Alignment Scores:
Pred. No.: 2.68e-63 Length: 156
Score: 819.00 Matches: 153
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 64.74% Indels: 0
DB: 2 Gaps: 0

US-10-728-051-2 (1-171) x Q941R0 (1-156)

QY 2 CTCACCATAGTAGCCCTCGCCCTTTCTCTCGTCCCGCCAGCATCTCGAGGCGAG 61
Db 4 LeuThrIleLeuValAlaLeuAlaLeuPheLeuLeuAlaHisAlaSerAlaArgGln 23
QY 62 CAGTGGGAACCTCCAAAGAGACAGAATGCCAGGCGAGCTCCAGAGGGCGAACCTGAGG 121
Db 24 GlnTrpGluLeuGlnGlyAspArgArgCysGlnSerGlnLeuGluArgAlaAsnLeuArg 43
QY 122 CCTGCGAGCAATCTCATGCGAGAGATCCAAACGTCGAGGAGATTTCATATGAACGGGAC 181
Db 44 ProCysGluGlnHisLeuMetGlnLysIleGlnArgAspGluAspSerTyrGluArgAsp 63
QY 182 CCGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCATATCATCGGAGGCGCTGGA 241
Db 64 ProTyrSerProSerGlnAspProTyrSerProSerProTyrAspArgGlyAlaGly 83

```



```

QY 242 TCCTCTCAGACCAAGAGAGGTTTGCATGAGCTGAAACGAGTTTGAGAACACCAAGG 301
Db 84 SerSerGlnHisGlnGluArgCysCysAsnGluLeuAsnGluPheGluAsnGlnArg 103
QY 302 TGCATGTGCAGGACATTCGCAACAGATCATGAGAACCCAGAGCGATAGTTGCAGGGGAGG 361
Db 104 CysMetCysGluAlaLeuGlnGlnLeuMetGluAsnGlnSerAspArgLeuGlnGlyArg 123
QY 362 CAACAGGAGCAACAGTTCAAGAGGAGCTCAGGAACCTTGCTCAACAGATGCGGCTTAGG 421
Db 124 GlnGlnGluGlnGlnPheGlyArgGluLeuArgAsnLeuProGlnGlnCysGlyLeuArg 143
QY 422 GCACCACAGGCTTGCAGCTTGGAGCTGCGAAAGTGGGGC 460
Db 144 AlaProGlnArgCysAspLeuAspValGluSerGlyGly 156

RESULT 4
Q8GV20 PRELIMINARY; PRT; 172 AA.
AC Q8GV20;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Allergen Ara h 2.02.
OS Arachis hypogaea (Peanut).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;
OC Arachis.
OX NCBI_TaxID=3818;
RN [1]
RP SEQUENCE FROM N.A.
RA Chatel J.-M., Bernard H., Orson F.M.;
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR ENBL: AY158467; AAN7576.1; -.
DR InterPro: IPR003612; AAI.
DR Pfam: PF00234; Tryp_alpha_aml1; 1.
DR SMART: SM00499; AAI; 1.
SQ SEQUENCE 172 AA; 20114 MW; B8BB91C8D8C143AB CRC64;

Alignment Scores:
Pred. No.: 6.04e-63 Length: 172
Score: 815.00 Matches: 155
Percent Similarity: 92.31% Conservative: 1
Best Local Similarity: 91.72% Mismatches: 1
Query Match: 64.43% Indels: 12
DB: 2 Gaps: 1

US-10-728-051-2 (1-717) x Q8GV20 (1-172)
QY 2 CTACCACTACTAGTAGCCCTCGCCCTTTCTCTCTGCTGCCACGCGATCTGCGAGGAG 61
Db 4 LeuThrIleLeuValAlaLeuAlaLeuPheLeuLeuAlaHisAlaSerAlaArgGln 23
QY 62 CAGTGGGAACCTCCAGGAGACAGAGATGCCAGAGCCAGCTCGAGAGGGCGAACCTGAGG 121
Db 24 GlnTrpGluLeuGlnGlyAspArgCysGlnSerGlnLeuGluArgAlaAsnLeuArg 43
QY 122 CCCTGCCAGCAGACATCTCATGCGAGAGATCCAACTGACGAGGATTCATATGACGGGAC 181
Db 44 ProCysGluGlnHisLeuMetGlnLysIleGlnArgAspGluAspSerTyrGlyArgAsp 63
QY 182 CCGTACAGCCCTAGCTCAG-----GATCCG 205
Db 64 ProTyrSerProSerGlnAspProTyrSerProSerGlnAspProAspArgAspPro 83
QY 206 TACAGCCCTAGTCCATATGATCGGAGAGCGCTGGATCCTCTCAGCACCAAGAGAGGTGT 265
Db 84 TyrSerProSerProTyrAspArgGlyAlaGlySerSerGlnHisGlnGluArgCys 103
QY 266 TGCATAGCTGAACGAGTTTGAGAACCAACAAAGGTGTCATGTCGAGGAGCATTCGACACAG 325
Db 104 CysAsnGluLeuAsnGluPheGluAsnGlnArgCysMetCysGluAlaLeuGlnGln 123

```

```

QY 326 ATCATGGAGAACCAAGAGCGATAGTTGCAGGGGAGGCAACAGAGCAACAGTTCAAGAGG 385
Db 124 IleMetGluAsnGlnSerAspArgLeuGlnGlyArgGlnGlnGlnGlnPheLysArg 143
QY 386 GAGCTCAGGAACCTTGCTCAACAGTCGCGCTTAGGGCCACAGGCTTGCAGCTTGGAC 445
Db 144 GluLeuArgAsnLeuProGlnGlnCysGlyLeuArgAlaProGlnArgCysAspLeuGlu 163
QY 446 GTCGAAAGTGGCGGCGAGACAGATAC 472
Db 164 ValGluSerGlyGlyArgAspArgTyr 172

RESULT 5
Q6PSU2 PRELIMINARY; PRT; 187 AA.
AC Q6PSU2;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Seed storage protein SSPI (Fragment).
OS Arachis hypogaea (Peanut).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;
OC Arachis.
OX NCBI_TaxID=3818;
RN [1]
RP SEQUENCE FROM N.A.
RA Yan Y., Wang L., Liao B., Li H., Lin X., Huang S.;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR ENBL: AY581853; AAR00598.1; -.
DR InterPro: IPR003612; AAI.
DR Pfam: PF00234; Tryp_alpha_aml1; 1.
DR SMART: SM00499; AAI; 1.
DR NON_TER 1; 1.
SQ SEQUENCE 187 AA; 21786 MW; 850AF22C94983FB8 CRC64;

Alignment Scores:
Pred. No.: 6.07e-63 Length: 187
Score: 815.00 Matches: 155
Percent Similarity: 92.31% Conservative: 1
Best Local Similarity: 91.72% Mismatches: 1
Query Match: 64.43% Indels: 12
DB: 2 Gaps: 1

US-10-728-051-2 (1-717) x Q6PSU2 (1-187)
QY 2 CTACCACTACTAGTAGCCCTCGCCCTTTCTCTCTGCTGCCACGCGATCTGCGAGGAG 61
Db 19 LeuThrIleLeuValAlaLeuAlaLeuPheLeuLeuAlaHisAlaSerAlaArgGln 38
QY 62 CAGTGGGAACCTCCAGGAGACAGAGATGCCAGAGCCAGCTCGAGAGGGCGAACCTGAGG 121
Db 39 GlnTrpGluLeuGlnGlyAspArgCysGlnSerGlnLeuGluArgAlaAsnLeuArg 58
QY 122 CCCTGCCAGCAACATCTCATGCGAGAGATCCAACTGACGAGGATTCATATGACGGGAC 181
Db 59 ProCysGluGlnHisLeuMetGlnLysIleGlnArgAspGluAspSerTyrGlyArgAsp 78
QY 182 CCGTACAGCCCTAGCTCAG-----GATCCG 205
Db 79 ProTyrSerProSerGlnAspProTyrSerProSerGlnAspProAspArgAspPro 98
QY 206 TACAGCCCTAGTCCATATGATCGGAGAGCGCTGGATCCTCTCAGCACCAAGAGAGGTGT 265
Db 99 TyrSerProSerProTyrAspArgGlyAlaGlySerSerGlnHisGlnGluArgCys 118
QY 266 TGCATAGCTGAACGAGTTTGAGAACCAACAAAGGTGTCATGTCGAGGAGCATTCGACACAG 325
Db 119 CysAsnGluLeuAsnGluPheGluAsnGlnArgCysMetCysGluAlaLeuGlnGln 138
QY 326 ATCATGGAGAACCAAGAGCGATAGTTGCAGGGGAGGCAACAGGAGCAACAGTTCAAGAGG 385

```


Alignment Scores: 2.53e-30 Length: 144
Pred. No.:

US-10-728-051-2 (1-717) x O8W251 (1-144)

RESULT 10

Q9SQG5	Q9SQG5	PRT;	129 AA.
ID	Q9SQG5	PRELIMINARY;	

DT	01-MAY-2000	(TrEMBLrel. 13, Created)
DI	01-MAY-2000	(TrEMBLrel. 13, Last sequence update)
DI	01-OCT-2002	(TrEMBLrel. 22, Last annotation update)
DE	Allergen.	
GN	Name=ArA h 7;	
OS	Arachis hypogaea (Peanut).	
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;	
OC	eurosid I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;	
OC	Arachis.	
RN	NCBI_TaxID=3818;	
RX	[1]	
RP	SEQUENCE FROM N.A.	
RC	SPRAIN=Virginia; TISSUE=Seed;	
RX	MEDLINE=99406463; PubMed=10474031; DOI=10.1159/000024203;	
RA	Kleber-Janke T., Cramer R., Appenzeller U., Schlaak M., Becker W.M.;	
RT	"Selective cloning of peanut allergens, including profilin and 2S	
RL	albumins, by phage display technology."	
RL	Int. Arch. Allergy Immunol. 119:265-274 (1999).	
DR	EMBL: AF091737; RAD56719.1; -.	
DR	InterPro; IPR003612; AA1.	
DR	SMART; SMO0499; AA1; 1.	
SR	SWISSPROT; P04499; AA1; 1.	
SQ	SEQUENCE 160 AA; 18417 MW; 9F9E4CEE6808D4C CRC64;	

Alignment Scores:

Pred. No.:	1.02e-17	Length:	160
Score:	300.00	Matches:	72
Percent Similarity:	61.82%	Conservative:	30
Best Local Similarity:	43.64%	Mismatches:	33
Query Match:	23.72%	Indels:	31
DB:	2	Gaps:	7

US-10-728-051-2 (1-717) x Q9SQH1 (1-160)

QY	2	CTCACACTACTAGTAGCCCTC-----GCCCTTTTCCTCTCGTGCC-----	43
Db	5	LeuSerIleLeuValAlaLeuLeuGlyAlaLeuLeuValValAlaSerAlaThrArgTrp 24	
QY	44	-----CACGCATCTCGCAGGCAGCAGTGGCAACTC-----CAAGGACGACAGA 88	
Db	25	AspProAspArgGlySerArgGlySerArgTrpAspAlaProSerArgGlyAspAspGln 44	
QY	89	TGCCAGAGCCAGCTCGAGAGGGCGCAACTGTAGGCCCTCGCAGCAACATCTCATGCAGAAG 148	
Db	45	CysGlnArgGlnLeuGlnArgAlaasnLeuArgProCysGluGluHisMetArgArgArg 64	
QY	149	ATCCAACTGACGAGGATTTCATATGAACGGACCCTGACAGCCCTAGTCAGGATCCGTAC 208	
Db	65	VaIGluGlnGluGlnGluGln---GluGlnAspGluTrp----- 76	
QY	209	AGCCCTAGTCCATATGATCGGAGAGGGCGCTGGATCCTCTCTCAG----- 250	
Db	77	-----ProTySerArgGlySerArgGlyArgGlnProGlyGluSerAspGlu 93	
QY	251	---CACCAGAGAGGTGTGCAATGACCTGAACAGATTTCAGAACCAACAAAGTGCATG 307	
Db	94	AenGlnGluGlnArgCysCysAsnGluLeuAsnArgPheGlnAsnGlnArgCysMet 113	
QY	308	TCCGAGGCATTTCACAACAGATCATCGAGAACACAGAGCGATAG----GTTGCAGGGAGGCA 363	
Db	114	CysGlnAlaLeuGlnGlnIleLeuGlnAsnGlnSerPheTrpValProAlaGlyGlnGlu 133	
QY	364	ACAGAGCAACAGTTCTAAGAGGGAGCTCAGAACTTGCTTCACACAGTGCAGGCTT 423	
Db	134	ProValAlaSerAspGlyGluGlyAlaaglnGluLeuAlaProgluLeuArgValGlnVal 153	
QY	424	ACCACAGCGTTGCCGA 438	
Db	154	ThrllysProLeuArg 158	

RESULT 12
Q647G8 PRELIMINARY; PRT; 158 AA.
ID Q647G8

```
AC Q647G8;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE 2S protein 2.
OS Arachis hypogaea (Peanut).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;
OC Arachis.
OX NCBI_TaxID=3818;
RN [1]
RP SEQUENCE FROM N.A.
RA Yan Y., Wang L., Huang S.;
RT "CDNA clone of peanut seed storage protein gene.";
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY72691; AAU21496.1; -.
SQ SEQUENCE 158 AA; 18455 MW; 32AD4D4061544948 CRC64;

Alignment Scores:
Pred. No.: 1.25e-17 Length: 158
Score: 299.00 Matches: 75
Percent Similarity: 63.12% Conservative: 26
Best Local Similarity: 46.88% Mismatches: 39
Query Match: 23.64% Indels: 21
DB: 2 Gaps: 7

US-10-728-051-2 (1-717) x Q647G8 (1-158)
QY 2 CTCACCATCTAGTACGCCCTC-----GCCCTTTTCTCTCTCTCGTCC----- 43
Db 4 LeuSerIleuValAlaLeuLeuGlyAlaLeuValAlaSerAlaThrArgTrp 23
QY 44 -----CACGCATCTCGGAGCGACGACGAGTGGGAACCTC-----CAAGGAGACAGAGA 88
Db 24 AspProAspArgGlySerArgGlyLeuArgTrpAspAlaProSerArgGlyAspAspGln 43
QY 89 TGCAGAGCCAGCTCGAGAGGGCGAACCTGAGGCCCTCGAGCAACATCTCATGCGAAG 148
Db 44 CysGlnArgGlnLeuGlnArgAlaAsnLeuArgProCysGlnGluHisIleArgGlnArg 63
QY 149 ATCCACGTCAGGAGGATTCATATGACGGGACCGGTACAGCCCTAGTCAGGATCCGTAC 208
Db 64 ValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 77
QY 209 AGCCCT-----AGTCCATATGATCGGAGAGCGCGTGGATCCTCTCAGCACCAAGAGAG 262
Db 78 SerGlnArgGlySerArgGlyArgArgProGlyGluSerAspGluAspGlnGlnArg 97
QY 263 TGTTCATAGCTGAACGAGTTTGAGAACCAACAAAGGTGCATGTGCGAGGCGATTGCAA 322
Db 98 CysCysAsnGluLeuAsnArgPheGlnAsnAsnGlnArgCysMetCysGlnAlaLeuGln 117
QY 323 CAGATCATGAGACCAACGAGGATAGTTG-----CAGGGAGGCAACAGGAGCAACAGTT- 378
Db 118 GlnIleuLeuAsnGlnSerPheArgPheGlnGlnAspArgSerGlnLeuHisGlnAsn 137
QY 379 CAAGAGGAGCTCAGGAACCTGCTCAACAGTGGCGGCTTAGGGCACACAGCGTTCGCA 438
Db 138 GlyGluGlyAlaGlnGluLeuAlaProGluLeuArgValGlnValThrLysProLeuArg 157

RESULT 13
Q99235 PRELIMINARY; PRT; 153 AA.
AC Q99235;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Conglutin delta.
OS Lupinus angustifolius (Narrow-leaved blue lupine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Genisteae; Lupinus.
```

```
OX NCBI_TaxID=3871;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=unicrop; Tissu=Cotyledon;
RX MEDLINE=91355912; PubMed=2103479;
RA Gayler K.R., Koliwas S., Macfarlane A.J., Lilley G.G., Baldi M.,
RA Blagrove R.J., Johnson E.D.;
RT "Biosynthesis, cDNA and amino acid sequences of a precursor of
RT conglutin delta, a sulphur-rich protein from Lupinus angustifolius.";
RL Plant Mol. Biol. 15:879-893(1990).
DR EMBL; X53523; CAA37598.1; -.
DR PIR; S12404; A33090.
DR InterPro; IPR003612; AAI.
DR Pfam; PF00234; Tryp_alpha_amy1; 1.
DR SMART; SM00499; AAI; 1.
SQ SEQUENCE 153 AA; 17785 MW; A45B7025353F0769 CRC64;

Alignment Scores:
Pred. No.: 8.7e-16 Length: 153
Score: 278.00 Matches: 62
Percent Similarity: 55.00% Conservative: 26
Best Local Similarity: 38.75% Mismatches: 46
Query Match: 21.98% Indels: 26
DB: 2 Gaps: 5

US-10-728-051-2 (1-717) x Q99235 (1-153)
QY 2 CTCACCATCTAGTACGCCCTC-----GCCCTTTTCTCTCTCTCGTCCCGCAGG 58
Db 4 LeuThrIleuLeuAlaLeuValAlaLeuValAlaValHisThrSerAlaPhe 23
QY 59 CAGCAGTGGGAACCTCCAGGAGACAGAGATGCCAGGCGAGCTCGAGGGCGCAACCTG 118
Db 24 Gln-----SerSerLysGlnSerCysLysArgGlnLeuGlnValAsnLeu 39
QY 119 AGCCCTCGGAGCAACATCTCATGACAGATCCACGTGACGAGGATTCATATGACGG 178
Db 40 ArgHisCysGluAsnHisIleAlaGlnArgIleGlnGlnGlnGlnGluGluAsp 59
QY 179 GACCCGTACAGC-----CCTAGTCAGGATCCG 205
Db 60 HisAlaLeuLysLeuArgGlyIleLysHisValIleLeuArgHisArgSerGlnGlu 79
QY 206 TACAGCCCTAGTCCATATGATCGGAGAGCGCTGGATCCTCTCAGCACCAAGAGGTGT 265
Db 80 TyrSerGlu-----GluSerGluGluLeuAspGlnCys 90
QY 266 TGCATGAGCTGACAGGTTTGAGAACCAACCAAGGTGCATGTGCGAGGCGATTGCAACAG 325
Db 91 CysGluGlnLeuAsnGluLeu---AsnSerGlnArgCysGlnCysArgAlaLeuGlnGln 109
QY 326 ATCATGAGAACCAAGAGCGATAGTTTGCAGGGAGGAGCAACAGGAGCAACAGGAGG 385
Db 110 IleTyrGluSerGlnSerGlnGlnCysGluGlySerGlnGlnGlnGlnGlnGlnGln 129
QY 386 GAGCTCAGGAACCTGCTCAACAGTCCGCGCTTAGGGCACACAGCGTTCGCGATTGGAC 445
Db 130 GluLeuGluLysLeuProArgThrCysGlyPheGlyProLeuArgCysAspValAsn 149

RESULT 14
2SS SOYBN STANDARD; PRT; 158 AA.
ID 2SS SOYBN
AC P19594;
DT 01-FEB-1991 (Rel. 17, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE 2S albumin precursor (GM2S-1) [contains: 2S albumin small chain
DE (Aspartic acid-rich peptide); 2S albumin large chain (8 kDa
DE methionine-rich protein) (8 kDa MRP)].
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
```

```

OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Hodgson 78; TISSUE=Cotyledon;
RA Galvez A.F., Revilla M.J.R., de Lumen B.O.;
RT "A novel methionine-rich protein from soybean cotyledon: cloning and
RL characterization of cDNA.";
RL (er) Plant Gene Register PGR97-103.
RN [2]
RP SEQUENCE OF 22-64.
RC TISSUE=Seed;
RX MEDLINE=87280104; PubMed=3611081;
RA Odani S., Koide T., Ono T.;
RT "Amino acid sequence of a soybean (Glycine max) seed polypeptide
RL having a poly(L-aspartic acid) structure.";
RN [3]
RP SEQUENCE OF 82-96.
RA Revilla M.J., Galvez A.F., Krenz D.C., de Lumen B.O.;
RT "An 8 kDa methionine-rich protein (MRP) from soybean (Glycine max)
RL cotyledon: identification, purification and N-terminal sequence.";
RN [4]
RP J. Agric. Food Chem. 44:2930-2935(1996).
CC -!- FUNCTION: This is a 2S seed storage protein.
CC -!- SUBUNIT: The protein consists of two chains linked by 2 disulfide
CC bonds.
CC -!- SIMILARITY: Belongs to the 2S seed storage albumins family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF005030; AAB71140.1; -.
DR PIR; T05710; T05710.
DR InterPro; IPR003612; AAI.
DR Pfam; PF00234; Tryp_alpha_aml1; 1.
DR SMART; SM00499; AAI; 1.
DR Direct protein sequencing; Seed storage protein; Signal.
KW SIGNAL
FT CHAIN 1 21
FT PROPEP 22 64 2S albumin small chain.
FT SITE 65 81
FT CHAIN 82 158 2S albumin large chain.
FT SITE 54 56 Cell attachment site (Potential).
FT DOMAIN 56 64 Poly-Asp.
FT DOMAIN 88 91 Poly-Glu.
SQ SEQUENCE 158 AA; 18460 MW; C1F42723B0F62D81 CRC64;

Alignment Scores:
Pred. No.: 1.03e-12 Length: 158
Score: 243.00 Matches: 56
Percent Similarity: 56.49% Conservative: 31
Best Local Similarity: 36.36% Mismatches: 55
Query Match: 19.21% Indels: 12
DB: 1 Gaps: 6

US-10-728-051-2 (1-717) x 2SS_SOYBN (1-158)
QY 5 ACCATACTAGTACCGCTTCCTCGCTGCCAGCATCTCGAGGCGAGCAG 64
DB 5 ThrLeuLeuLeuSerLeuLeuPheCysIle---AlaHisThrCysSerAlaSerLys 23
QY 65 TGGAACTCCAGAGACAGAGAGATCCAGAGCCAGCTCGAGAGGCGGCACTGAGGCC 124
DB 24 TrpGlnHisGlnAsp---SerCysArgLysGlnLeuGlnGlyValAsnLeuThrPro 42
QY 125 TGGAGCAACATCTTCATGCAGAGATCCAA-----CGTGACGAGATTCATAGACGG 178
DB 43 CysGlnHisLeuMetGlnGlyIleGlnGlyArgGlyAspAspAspAsp 62
QY 179 GACCCGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCATATATATGTCGAGAGCGCT 238

```

```

Db 63 AspAspAsnHisIleLeuArgThrMetArgGlyArgIleAsnTyrIleArgAsnGlu 82
QY 239 GGATCTCTCAG-----CACCAGAGAGAGTGTGTGCAATGAGCTGAAC 280
Db 83 GlyLysAspGluAspGluGluGluGlyHisMetGlnLysCysCysThrGluMetSer 102
QY 281 GAGTTTGAAGAACCAACCAAGGTGTCATGTCGAGGCAATTCACAGATCATGAGAACCA 340
Db 103 GluLeu---ArgSerProLysCysGlnCysLysAlaLeuGlnLysIleMetGluAsnGln 121
QY 341 AGCATATAGTTTGCAGGGGAGGCAACAGAGCAACAGTTCACAGAGGAGCTCAGGAACCTG 400
Db 122 SerGluGluLeuGlu---GluLysGlnLysLysLysMetGluLysGluLeuLeuLeu 140
QY 401 CCTCAACAGTGGCGCTTAGGGCACCAACAGCGTTCGACTTG 442
Db 141 AlaThrMetCysArgPheGlyProMetIleGlnCysAspLeu 154

RESULT 15
Q9ZNZ4
ID Q9ZNZ4 PRELIMINARY; PRT; 155 AA.
AC Q9ZNZ4;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Napin-type 2S albumin 1 precursor.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=variety PIONEER 9341; TISSUE=Seed;
RA Jung R., Hastings C., Coughlan S.J., Hu W.-N.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U71194; AAD09630.1; -.
DR InterPro; IPR003612; AAI.
DR Pfam; PF00234; Tryp_alpha_aml1; 1.
DR SMART; SM00499; AAI; 1.
KW Signal.
FT SIGNAL 1 20 Potential.
FT CHAIN 21 155 napin-type 2S albumin 1.
SQ SEQUENCE 155 AA; 17835 MW; 2C15E55E936301E3 CRC64;

Alignment Scores:
Pred. No.: 1.4e-12 Length: 155
Score: 241.50 Matches: 58
Percent Similarity: 57.14% Conservative: 30
Best Local Similarity: 37.66% Mismatches: 53
Query Match: 19.09% Indels: 13
DB: 2 Gaps: 7

US-10-728-051-2 (1-717) x Q9ZNZ4 (1-155)
QY 2 CTCACATACTAGTACCGCTTCCTCGCTGCCAGCATCTCGAGGCGAG 61
DB 4 LeuThrIleLeuLeuAlaLeuLeuPheIle-----AlaHisThrCysCysAlaSer 21
QY 62 CAGTGGGAACCTCCAGGAGAGAGATCCAGAGCCAGCTCGAGAGGCGGCACTGAGG 121
DB 22 LysTrpGlnGlnHisGlnGlnGlnSerCysArgGluGlnLeuLysGlyIleAsnLeuAsn 41
QY 122 CCCTCGAGAGCAACATCTCATGCAGAGATCCAA-----CGTGACGAGATTCATAT 172
DB 42 ProCysGlu---HisIleMetGluLysIleGlnAlaGlyArgArgGlyGluAspGlySer 60
QY 173 GAACGGGACCGGTACAGCCCTAGTCAGATCCGTACAGCCCTAGTCATGATGTCGAGA 232
DB 61 AspGluAspHisIleLeuLeuLeuLeuArgThrMetPro---GlyArgIleAsnTyrIleArgLys 79
QY 233 GGCCTGATCCTCTCAG-----CACCAGAGAGAGTGTTCGATGAGCTGNAC 280

```

```

Db      80 LysGluGlyLysGluGluGluGluGluGlyHisMetGlnLysCysCysSerGluMetSer 99
QY      281 GAGTTTGAGAACCAACCAAGTGCATGTGCGAGGCATTGCAACAGATCATGGAGAACCCAG 340
Db      100 GluLeuLys--SerProIleCysGlnCysLysAlaLeuGlnLysIleMetAspAsnGln 118
QY      341 AGCGATAGTTGCGAGGGAGGCAACAGGAGCAACAGTTCAAGAGGGAGCTCAGGAACCTTG 400
Db      119 SerGluGlnLeuGluGlyLys--GluLysLysGlnMetGluArgGluLeuMetAsnLeu 137
QY      401 CCTCAACAGTGGCGCCTTAGGGCACACACAGCGTTGGGACTTG 442
Db      138 AlaIleArgCysArgLeuGlyProMetIleGlyCysAspLeu 151

```

Search completed: April 8, 2005, 05:25:11 ~
Job time : 128.066 secs

This Page Blank (uspto)

43 denazrIeIuSewIwGlvphelGlvthrtmAnpPrAgnAnIvAprophe 62

```

QY 121 GAATGCCCGCGCTGCGCTCTCTCGCTTAGTCTCCGCGCAACGCGCTTCGTAGCGCT 180
Db 63 GlnCysAlaGlyValAlaLeuSerArgCysThrLeuAsnArgAsnAlaLeuArgPro 82
QY 181 TTCTACTCCATGCTCCCGAGGAGATCTTCATCAGCAAGGAGGAGTACTTTGGGTG 240
Db 83 SerTyrThrAsnAlaProGlnGluIleTyrGlnGlnGlySerGlyIlePheGlyMet 102
QY 241 ATATTCTCTGTTGCTCTAGACTATGAAGAGCTCACACACAAGTCTGCTGATCTCAG 300
Db 103 IlePheProGlyCysProSerThrPheGluGluProGlnGlnGlyGly-----Gln 119
QY 301 TCCCAAGACCAACAAGAGCTCTCCAAAGGAGAAACCAAGCAACAGCAACGAGATAGT 360
Db 120 SerSerArgPro-----GlnAspArg 126
QY 361 CACCAAGAGTGCACCGTTTCGATGAGGTGATCTATTGCACTTCCACCGGTGTGCT 420
Db 127 HisGlnLysIleTyrHisPheArgGluGluIleValProThrGlyPheAla 146
QY 421 TTCTGGCTCTACCAACGACCAACGACACTGATGTTGTTGCTGTTCTTCTACTGACACCAAC 480
Db 147 TyrTipMetTyrAsnAsnGluAspThrProValValAlaValSerLeuIleAspThrAsn 166
QY 481 AACACGACACACAGCTTGATGATGATCCCGAGAGATTCATTTGCTGGGAAACAGGAG 540
Db 167 SerPheGlnAsnGlnLeuAspGlnMetProArgArgPheTyrLeuAlaGlyAsnGlnGlu 186
QY 541 CAAGAGTCTTAAGTACGACCAACAAAGCAGACAGAGCAGAGCAGAGAGCTTACCATAT 600
Db 187 GlnGluPheLeuGlnTyrGlnProGlnLysGlnGlnGly-----199
QY 601 AGCCCATACAGCCCGCAAGTACAGCTAGACAGAGAGCGTGAATTTAGCCCTCGAGAG 660
Db 200 -----Gly 200
QY 661 CACCAAGCGCGCAGAGAACGAGCAGACAGAGAGAGAGAAAGCAAGGTGGAACATCTTC 720
Db 201 ThrGlnSerGlnLysGlyLysArgGlnGlnGluGluGluAsnGluGlyGlySerIleLeu 220
QY 721 AGCGGTCTACGCGCGAGTCTCCGAAACAGCCTCCAGGTTGACACAGACAGATAGT 780
Db 221 SerGlyPheAlaProGluPheLeuGluHisAlaPheValVal-----AspArgGlnIleVal 239
QY 781 CAAACCTTAAGCGCAGACCGAGAGTGAAGAGAGGAGCGCATTTGACAGTGGGGA 840
Db 240 ArgLysLeuGlnGlyGluAsnGluGluGluGlyValAlaIleValThrValLysGly 259
QY 841 GGCCTCAGATCTTGAGCCCA-----GATAGAAAGAGAGCTGCCAGCAAGAGAG 891
Db 260 GlyLeuSerValIleSerProThrGluGluGlnGlnArgProGluGluGluGlu 279
QY 892 GAATACGATGAAGATGAATATGATACGATGAAGAGATAGAGGCGTGGCAGGGAAGC 951
Db 280 LysProAsp-----CysAspGluLysAspLysHisCysGlnSerGlnSer 294
QY 952 AGAGCAGGCGGAATGGTATTGAAGAGACGATCTGCACCGCAAGTCTAAAGAACATP 1011
Db 295 Arg-----AsnGlyIleAspGluThrIleCysThrMetArgLeuArgHisAsnIle 311
QY 1012 GTTAGAAACAGATCCCTGACATCTACACCTCAGCTGTTTCACTCAAACTGCCAAC 1071
Db 312 GlyIleThrSerSerProAspIlePheAsnProGlnAlaGlySerIleThrAlaThr 331
QY 1072 GATCTCAACCTCTTAATCTAGTGGCTTGGACCTAGTGTGAATATGAAATCTCTAC 1131
Db 332 SerLeuAspPheProAlaLeuSerTrpLeuLysLeuSerAlaGlnPheGlySerLeuArg 351
QY 1132 AGAATGATGATTTGTTGCTGCTACTACACCAACGACGACAGCATCATATATCATG 1191
Db 352 LysAsnAlaMetPheValProHisTyrAsnLeuAsnAlaAsnSerIleIleTyrAlaLeu 371
QY 1192 AGGGGACGGGCTCAGCTGCAAGTCGTGGAGACGACCAAGCGGCAACAGAGTGTACGACGGAG 1251

```

```

Db 372 AsnGlyArgAlaLeuValGlnValValAsnCysAsnGlyGluArgValPheAspGlyGlu 391
QY 1252 CTTCAAGAGGTCACGTGCTTGTGGTGCACAGAACTTCGCGCTCGCTGGAAAGTCCAG 1311
Db 392 LeuGlnGluGlyGlnValLeuIleValProGlnAsnPheAlaValAlaAalaargSerGln 411
QY 1312 ACGGAACTTCGAATACGTGTCATTCAGACAGACTCAAGGCCAGCATACCCACCTC 1371
Db 412 SerAspAsnPheLutyrValSerPheLysThrAsnAspArgProSerIleGlyAsnLeu 431
QY 1372 GCCGTGAAAACCTCCGTATAGTAACTCCGCGGAGGAGTGTGTCAAATTCATATGCG 1431
Db 432 AlaGlyAlaAsnSerLeuLeuAsnAlaLeuProGluGluValIleGlnGlnThrPheAsn 451
QY 1432 CTCCTCAAGGAGCAGGCAAGCAGCTTAAAGACACAAACCCCTTCAAGTCTTCGTGTC 1491
Db 452 LeuArgArgGlnGlnAlaArgGlnValLysAsnAsnProPheSerPheLeuValPro 471
QY 1492 CCGTCTCAGCAGCTCCGAGGCGTGTGGCT 1521
Db 472 ProLysGluSerGlnArgArgValValAla 481

```

RESULT 2

S11002

glycinin G2 precursor - soybean

N:Alternate names: glycinin A2B1a

C:Species: Glycine max (soybean)

C>Date: 21-Nov-1993 #sequence revision 19-Jan-1996 #text_change 09-Jul-2004

C:Accession: S11002; S04604; A26990

R:Nielsen, N.C.; Dickinson, C.D.; Cho, T.J.; Thanh, V.H.; Scallion, B.J.; Fischer, R.L.;

Plant Cell 1, 313-328, 1989

A:Title: Characterization of the glycinin gene family in soybean.

A:Reference number: S10851; MUID:92393391; PMID:2485233

A:Accession: S11002

A:Molecule type: DNA

A:Residues: 1-485 <NIE>

A:Cross-references: UNIPROT:P04405

A:Experimental source: variety Dare

R:Thanh, V.H.; Tumer, N.E.; Nielsen, N.C.

Nucleic Acids Res. 17, 4387, 1989

A:Title: The glycinin G2 gene from soybean.

A:Reference number: S04604; MUID:89296499; PMID:2740230

A:Molecule type: DNA

A:Residues: 1-485 <THA>

A:Cross-references: EMBL:X15122; NID:gl8636; PIDN:CAA33216.1; PID:gl8637

A:Experimental source: variety Dare

R:Fukazawa, C.; Momma, T.; Higuchi, W.; Uda, K.

Nucleic Acids Res. 15, 8117, 1987

A:Title: Complete nucleotide sequence of the gene encoding a glycinin A2B1a subunit prec

A:Reference number: A26990; MUID:88040439; PMID:3671077

A:Molecule type: DNA

A:Residues: 1-485 <FUK>

A:Cross-references: GB:X02806

C:Genetics:

A:Gene: Gv2

A:Introns: 91/1, 177/3, 356/3

C:Superfamily: glycinin

C:Keywords: storage protein

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-485/Product: glycinin G2 #status predicted <MAT>

Alignment Scores:

Pred. No.:	1-57e-103	Length:	485
Score:	1499.50	Matches:	297
Percent Similarity:	72.02%	Conservative:	71
Best Local Similarity:	58.12%	Mismatches:	94
Query Match:	55.72%	Indels:	49
DB:	2	Gaps:	6

US-10-728-051-3 (1-1524) x S11002 (1-485)

F;20-310/Product: glycinin Ala chain #status predicted <MAT2>
F;311-495/Product: glycinin Bx chain #status predicted <MAT3>

Alignment Scores: 2,87e-103 Length: 495
Pred. No.: 1496.00 Matches: 292
Score: 70.87% Conservative: 73
Best Local Similarity: 56.70% Mismatches: 100
Query Match: 55.59% Indels: 50
DB: 2 Gaps: 6

US-10-728-051-3 (1-1524) x S10851 (1-495)

QY 1 CGCAGCAACCGAGAGAGACGGCTCCAGTTCACCGCCTCAATGCCGAGAGACCTGAC 60
DB 23 ArgGluGlnProGlnAsnGluCysGlnLeuLysLeuAsnAlaLeuLysProAsp 42
QY 61 AATCGCATTAATCAGAGCGGCTTACATTGAGACTTGGAAACCCCAACACAGAGAGTTTC 120
DB 43 AsnArgileGluSerGluGlyGlyLeuileGluThrTrpAsnProAsnAsnLysProPhe 62
QY 121 GAATGGCGCGCGCTCTCTCTCGCTTAGTCTCCCGCGCAACGCCCTTCGTAGGCT 180
DB 63 GlnCysAlaGlyValAlaLeuSerArgCysThrLeuAsnArgAsnAlaLeuArgPro 82
QY 181 TTCTACTCAATGCTCCCGAGGAGATTCATCCAGCAAGAGGGGATCTTTGGTTG 240
DB 83 SerTyrThrAsnGlyProGlnGluileTyrileGlnGlnGlyLysGlylePheGlyMet 102
QY 241 ATATTCCTCGTGTCTCTAGCACTATGAAGAGCCTCACACAAAGTCTCGTATCTAG 300
DB 103 IleTyrProGlyCysProSerThrPheGluProGlnProGlnGlnArgGlyGln 122
QY 301 TCCCAAGACACCAAGAGCTCTCCAGGAGAGACCAAGCCAAAGCAAGAGATAGT 360
DB 123 SerSerArgPro-----GlnAspArg 129
QY 361 CACGAGAGGTGGACCGTTTCGATGAGGTGATCTCATTCAGTTCCTCCACCGGTGTGCT 420
DB 130 HisGlnlyleTyrAsnPheArgGluGlyAspLeuileAlaValProThrGlyValAla 149
QY 421 TTCTGGCTCTACACGACACGACGACTGATGTTGCTGTTCTCTTACTGACACMAC 480
DB 150 TrpTrpMetTyrAsnAsnGluAspThrProValValAlaValSerileleAspThrAsn 169
QY 481 AACACGACACACGCTGATGATCTCCAGGAGATTCATTTGGCTGGGAAACAGCGAG 540
DB 170 SerLeuGluAsnGlnLeuAspGlnMetProArgArgPheTyrLeuAlaGlyAsnGlnGlu 189
QY 541 CAAGAGTTCCTTAAGGTACCAACAAAGACGACAGACAAAGACGACGAGAGCTTACCATAT 600
DB 190 GlnGluPheLeuLysTyrGlnGlnGln----- 199
QY 601 AGCCCATACAGCCCAAGTCCAGCTAGACAGAGAGCGTGAATTTAGCCCTCGAGGA 660
DB 200 -----GlyGly 201
QY 661 CAGCAGCGCGCAGAGACGAGCAGGACAAAGAAAGAAAGAGTGGAAACATCTTC 720
DB 202 HisGlnSerGlnlyleGlyHisGlnGlnGlnGluGluAsnGluGlySerileleu 221
QY 721 AGCGGCTTCACCGGAGTTCCTGGAACAGACCTTCAGGTTGACGACGACAGATAGTG 780
DB 222 SerGlyPheThrLeuGluPheLeuGluHisAlaPheSerVal---AspLysGlnleAla 240
QY 781 CAAAACCTTAAGCGGACGACGAGAGTGAAGAGAGGAGGAGCTTGTACAGTGGAGGGA 840
DB 241 LysAsnLeuGlnGlyGluAsnGluGlyGluAspLysGlyAlaileValThrVallysgly 260
QY 841 GGCCTCAGATCTTGAGCCCA-----GATGAAAGACAGACGTGCGCAGAGAGAGAG 891
DB 261 GlyLeuSerValileLysProThrAspGluGlnGlnArgProGlnGlnGluGlu 280

QY 892 GAATACGATGAAGATGAATGAATATAC-----GATCAAGAG---GATAGAAGG 936
DB 281 GluGluGluGluAspGluLysProGlnCysLysGlyLysAspLysHisCysGlnArgPro 300
QY 937 CGTGGCAGGGAAGCAGAGCAGGCGGGAATGTTATTGAAGAGACGATCTGCACCGCAAGT 996
DB 301 ArgGlySerGlnSerLysSerArgArgAsnGlyleAspGluThrileCysThrMetArg 320
QY 997 GCTAAAAAGAACATTGGTAGAAACAGATCCCTTGACATCTTACAAACCTCAAGTGTGTTCA 1056
DB 321 LeuArgHisAsnleGlyGlnThrSerSerProAspIleTyrAsnProGlnAlaGlySer 340
QY 1057 CTCAAACTGCCAAGCATCTCAACCTTCTAATTAATAGTGGCTTGGACCTAGTGTGAA 1116
DB 341 ValThrThrAlaThrSerLeuAspPheProAlaLeuSerTrpLeuArgLeuSerAlaGlu 360
QY 1117 TATGGAATCTCTACAGGAATGATCTTGTTCGCTCACTACAAACACCAACGACACACAGC 1176
DB 361 PheGlySerLeuArgLysAsnAlaMetPheValProHisTyrAsnLeuAsnAlaAsnSer 380
QY 1177 ATCATATATCGATTGAGGCGCGCTCACGTGCAATCTGTGACACGACCAACGCAACAGA 1236
DB 381 IleileTyrAlaLeuAsnGlyArgAlaLeuileGlnValValAsnCysAsnGlyGluArg 400
QY 1237 GTGTACGAGGAGGACTTCAAGAGGTCAAGGCTGCTGCTGTCGACAGAACTTCGCCGTC 1296
DB 401 ValPheAspGlyGluLeuGlnGluGlyArgValleuileValProGlnAsnPheValVal 420
QY 1297 GCTGMAAGTCCAGAGCGAGAACTTCGAATACGTGTCATTCAGACAGACTCAAGGCC 1356
DB 421 AlaAlaArgSerGlnSerAspAsnPheGluTyrValSerPheLysThrAsnAspThrPro 440
QY 1357 AGCATAGCAACCTCGCGGTGAAATCTCGTCTCATATACTTCGCGGAGGAGTGTGTT 1416
DB 441 MetileGlyThrLeuAlaGlyAlaAsnSerLeuLeuAsnAlaLeuProGluGluValle 460
QY 1417 GCAATTCATATGGCTCCAAAGGAGCAGGACGAGCAAGGAGCTTAAAGAACAAACCCCTTC 1476
DB 461 GlnHisThrPheAsnLeuLysSerGlnGlnAlaArgGlnileLysAsnAsnAsnProPhe 480
QY 1477 AAGTTCCTTCGTTCCACCTCTCAGCAGCTCTCCGAGGGCTGTGGCT 1521
DB 481 LysPheLeuValProGlnGluSerGlnLysArgAlaValAla 495
RESULT 4
FWSYGI
glycinin chain A2Bla precursor - soybean
N;Alternate names: 11S globulin
N;Contains: glycinin chain A2; glycinin chain Bla
C;Species: Glycine max (soybean)
C;Date: 31-Mar-1988 #sequence revision 31-Mar-1988 #text change 09-Jul-2004
C;Accession: A91341; A92454; E92454; A90024; A92452; S10503; S74123; A05082; A05164; A051
R;Mamma, T.; Negro, T.; Uda, K.; Fukazawa, C.
FEBS Lett. 188, 117-122, 1985
A;Title: A complete cDNA coding for the sequence of glycinin A2Bla subunit precursor.
A;Reference number: A91341
A;Accession: A91341
A;Molecule type: mRNA
A;Residues: 1-485 <MOM>
A;Cross-references: UNIPROT:P04405
A;Experimental source: strain Bonimori
A;Note: the source of this mRNA was cotyledon tissue taken from seeds at the middle stage
P;Marco, Y.A.; Thanh, V.H.; Tumer, N.E.; Scallan, B.J.; Nielsen, N.C.
J. Biol. Chem. 259, 13436-13441, 1984
A;Title: Cloning and structural analysis of DNA encoding an A2Bla subunit of glycinin.
A;Reference number: A92454; MUID:85030472; PMID:6092376
A;Accession: A92454
A;Molecule type: mRNA
A;Residues: 262-446 <MAL>
A;Accession: B92454
A;Molecule type: DNA
A;Residues: 318-485 <MA2>
A;Experimental source: strain CX635-1-1-1

Qy	181	TTCTACTCCAAATGCTCCCCAGGAGATCTTTCATCCAGCAAGGAGGATACTTTGGGTG	240
Db	80	SerTyr-ThrAsnGlyProGlnGluIleTyrIleGlnGlnGlyAsnGlyIlePheGlyMet	99
Qy	241	ATATTCCCTGGTTGCTCTAGACACTATGAAGAGCGCTCACACAGAGTCTCGATCTCAG	300
Db	100	IlePheProGlyCysProSerThrTyrGlnGluProGlnGluSerGlnGlnArgGlyArg	119
Qy	301	TCCCAAGACCCCAAGACGCTCTCCAAGGAGAGAACCAAGCCAAACAGCAACGAGATAGT	360
Db	120	SerGlnArgPro-----GlnAspArg	125
Qy	361	CACCAGAAGGTGCACCGTTTCGATGAGGGTGATCTCATTCAGTTCCTCCACCGGTGTGCT	420
Db	127	HisGlnLysValHisArgPheArgGluGlyAspLeuIleAlaValProThrGlyValAla	146
Qy	421	TTCTGGCTCTCAACAGCAGCAGCAGCTGATGTTGTGTGCTTTCTCTTACTGACACCAAC	480
Db	147	TrpTrpMetTyrAsnAsnGlnAspThrProValValAlaValSerIleIleAspThrAsn	166
Qy	481	AACAAACGACAAACAGCTTGATCAGTTCCCCAGAGAGATTCAATTGGCTGGGAACACCGAG	540
Db	167	SerLeuGlnAsnGlnLeuAspGlnMetProArgArgPheTyrLeuAlaGlyAsnGlnGlu	186
Qy	541	CAAGAGTTCTTAAGGTACCAACCAACAAAGCAGACAAAGCAGACGAGAAAGCTTACCATAT	600
Db	187	GlnGluPheLeuLysTyrGlnGln-----	195
Qy	601	AGCCCATACAGCCCGCAAGTCAGCTAGACAGAGAGCGTGAAATTAGCCCTCGAGGA	660
Db	196	-----GlnGlnGlyGlySerGlnSerGlnLysGly	205
Qy	661	CAGCACAGCCGACAGAAACGAGCAGGACAGAAAGAAACCAAGGTGGAACATCTTC	720
Db	206	LysGln-----GlnGluGluGlnAsnGluGlySerAsnIleLeu	218
Qy	721	AGCGGCTTCACGCGGAGTTCTCTGGAACAGCGCTTCAGGTTCACGACACACAGATAGTG	780
Db	219	SerGlyPheAlaProGluPheLeuLysGluAlaPheGlyVal---AsnMetGlnIleVal	237
Qy	781	CAAAACCTAAGAGCGGAGACCGAGAGTGAAGAGAGGAGGCCATTGTGACAGTGAAGGGGA	840
Db	238	ArgAsnLeuGlnGlyGluAsnGluGluGluAspSerGlyAlaIleValThrValLysGly	257
Qy	841	GGCCTCAGATCTTGAGCCAGATAGAAAGAGACGTGCCGACGAGAGAGGAGGAATACGAT	900
Db	258	GlyLeuArgValThrAlaProAlaMetArgLysProGlnGlnGluAlaAspAspAspAsp	277
Qy	901	GAAGATGAA-----TATGAATACGATGAAGAGGATAGAAGCGCTGGCAGGGGA	948
Db	278	GluGluGluGlnProGlnCysValGluThrAspLysGlyCysGlnArgGlnSerLysArg	297
Qy	949	AGCAGAGCAGGGGGAAATGTATTGAAGAGACAGCATCTGCACCGCAAGTGTCTAAAGAAGAC	1008
Db	298	SerArg-----AsnGlyIleAspGluThrIleCysThrMetArgLeuArgGlnAsn	314
Qy	1009	ATTGGTAGAAACAGATCCCTCGACACTCAACCCCTCAAGCTGGTTCACCTCAAAACCTGCC	1068
Db	315	IleGlyGlnAsnSerProAspIleTyrAsnProGlnAlaGlySerIleThrThrAla	334
Qy	1069	AACGATCTCAACCTTCTAATCATAGTGCGCTGGACCTAGTGTCTGAATATGGAATCTC	1128
Db	335	ThrSerLeuAspPheProAlaLeuTyrPheLeuLysLeuSerAlaGlnTyrGlySerLeu	354
Qy	1129	TACAGGAATGCATTGTTGTGCTCACTACAACCAACCCACACAGCATCATATATCGA	1188
Db	355	ArgLysAsnAlaMetPheValProHisTyrThrLeuAsnAlaAsnSerIleIleTyrAla	374
Qy	1189	TTGAGGGGCGGCTCACTGTCGAAGTCTGTCGACGACGACCGGCAACAGAGTGTACGACGAG	1248
Db	375	LeuAsnGlyArgAlaLeuValGlnValValAsnCysAsnGlyGluArgValPheAspGly	394

Qy 1177 ATCATATATCGATTGAGGGGACGGGTCTACGTGCATAAGTCGTGTGAACAGCAACGCCAACAGA 1236
Db 381 lletltyrAlaLeuasnGlyArgAlaLeuileGlnValValAsnCysAenGlyGluArg 400
Qy 1237 GTGTACGACGAGGAGCTTCAAGAGGGTCACGTGCTTGTTGGTGGCCACAGAACTTTCGCCGTC 1296
Db 401 ValPheAspGlygluleuGlnGluGlyArgValLeuileValProGlnAsnPheValVal 420
Qy 1297 GCTGGAAGATCCCAGAGCGAGAACTTCGAATACGTGGCATTCAAAGACAGACTCAAGGCC 1356
Db 421 AlaAlaAArgSerGlnSerAspAsnPheGluTyValSerPhelystThrAsnAspThPro 440
Qy 1357 AGCATAGCCAACCTCGCGGTGAAAACCTCGTCATAGATAACCTCGCGAGGAGGTGGTT 1416
Db 441 MetlleGlyThrLeuAlaGlyAlaAsnSerLeuLeuAsnAlaLeuProGluGluValile 460
Qy 1417 GCAAATTCATATGGGCTCCAAAGGGAGCAGGCAAGGCAGCTTAAGAACACACACCCTTC 1476
Db 461 GlnHisThrPheAsnLeuLysSerGlnGlnAlaAArgGlnIleLysAsnAsnProPhe 480
Qy 1477 AAGTCTCTTCGTTCCACCGCTCCAGAGTCTCCGAGGGCTGTGGCT 1521
Db 481 LysPheLeuValProProGlnGlnSerGlnLysAArgAlaValala 495

RESULT 6
S44294
legumin A precursor - spring vetch
C:Species: Vicia sativa (spring vetch, tare)
C>Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 09-Jul-2004
C:Accession: S44294
R:Nong, V.; Becker, C.; Muentz, K.
submitted to the EMBL Data Library, April 1994
A:Description: Cloning and heterologous expression of cDNAs encoding legumins of vetch
A:Reference number: S44266
A:Accession: S44294
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-498 <NON>
A:Cross-references: UNIPROT:Q41702; EMBL:Z32835; NID:g483448; PIDN:CAA83677.1; PID:g483448
C:Superfamily: glycinin

Alignment Scores:

Pred. No.:	2,09e-100	Length:	498
Score:	1457.50	Matches:	289
Percent Similarity:	68.8%	Conservative:	74
Best Local Similarity:	54.84%	Mismatches:	93
Query Match:	54.16%	Indels:	71
DB:	2	Gaps:	8

US-10-728-051-3 (1-1524) x S44294 (1-498)

Qy	1 CGGAGCACACCGGAGGAGACCGGTGCCAGTCTCCAGCGCTCAATCGCAGACACCTGAC	60
Db <td>23 ArgGluGlnSerGlnGlnAsnGluCysGlnLeuGluArgIleAsnAlaLeuGluProAsp</td> <td>42</td>	23 ArgGluGlnSerGlnGlnAsnGluCysGlnLeuGluArgIleAsnAlaLeuGluProAsp	42
Qy <th>61 AATGCAATGTAATCAGAGCGCGTTACATTGAGACTTTGGAACCCCAACACAGGAGTTC</th> <th>120</th>	61 AATGCAATGTAATCAGAGCGCGTTACATTGAGACTTTGGAACCCCAACACAGGAGTTC	120
Db <th>43 AsnArgIleGluSerGluGlyLeuIleGluThrTrpAsnProAsnAsnAArgGlnPhe</th> <th>62</th>	43 AsnArgIleGluSerGluGlyLeuIleGluThrTrpAsnProAsnAsnAArgGlnPhe	62
Qy <th>121 GAATGCGCGCGGTGCGCCCTCTCTCGCTTAGTCTCCGCGCAACGCCCTTCGTAGGCCT</th> <th>180</th>	121 GAATGCGCGCGGTGCGCCCTCTCTCGCTTAGTCTCCGCGCAACGCCCTTCGTAGGCCT	180
Db <th>63 ArgCysAlaAArgValAlaLeuSerArgAlaThrLeuGlnArgAsnAlaLeuAArgPro</th> <th>82</th>	63 ArgCysAlaAArgValAlaLeuSerArgAlaThrLeuGlnArgAsnAlaLeuAArgPro	82
Qy <th>181 TTCTACTCCAATGCTCCCGAGAGATCTTCATCCAGCAAGGAGGGATACTTTGGGTTG</th> <th>240</th>	181 TTCTACTCCAATGCTCCCGAGAGATCTTCATCCAGCAAGGAGGGATACTTTGGGTTG	240
Db <th>83 TyrTrpSerAsnAlaproGlnGluIleTyrIleGlnGlnGlyAsnGlyTyrPheGlyMet</th> <th>102</th>	83 TyrTrpSerAsnAlaproGlnGluIleTyrIleGlnGlnGlyAsnGlyTyrPheGlyMet	102
Qy <th>241 ATATTCCCTGGTGTGCTAGACACTATGAAGAGCTTCACACACAAGTCGTCGATCTCAG</th> <th>300</th>	241 ATATTCCCTGGTGTGCTAGACACTATGAAGAGCTTCACACACAAGTCGTCGATCTCAG	300
Db <th>103 ValPheProGlyCysPProGluThrHisGluGluPro-----GlnGln</th> <th>116</th>	103 ValPheProGlyCysPProGluThrHisGluGluPro-----GlnGln	116
Qy <th>301 TCCCAAGACCAACCAAGAGCTCTCCAAAGGAGAGACCAAGCCAAAGCAACGACGAGTAGT</th> <th>360</th>	301 TCCCAAGACCAACCAAGAGCTCTCCAAAGGAGAGACCAAGCCAAAGCAACGACGAGTAGT	360

117	SerGlu-----GlnGlyGlu-----GlyArgArgTyrArgAspSer	128
361	CACCAGAAGGTGCACCGCTTTTCGATGAGGGTGATCTCATTCAGTTCACCGGTGGTGGCT	420
129	HisGlnIysValAsnArgPheArgGluGlyAspIleIleAlaValProThrGlyIleAla	148
421	TTCTGGCTCTPACAACGACCAACGACACTGATGTTGTTGCTGCTTTCTCTTACTGACACCAAC	480
149	PheTrpMetTyrAsnAspGlnAspThrProValIleAlaIleSerLeuThrAspThrGly	168
481	AACAAACGACACACAGCTTGATCACTGTTCCCGACGAGATTCATTTGGCTGGGACCGGAG	540
169	SerSerAsnGlnLeuAspGlnMetProArgArgPheTyrLeuAlaGlyAsnGlnGlu	188
541	CAAGAGTCTTTAAGGTACACGACCAACAAAGCAGACAAAGCAGACCAAGAAAGCTTTACCATAT	600
189	GlnGluPheLeuArgTyrGlnHisGln-----	197
601	AGCCCATACAGCCCGCAAAAGTCAGCTAGCAACAAGAGCGTGAAATTTAGCCCTCGAGGA	660
197	-----	197
661	CAGCACAGCCGACAGAGAACGACGACGACCAAGACAGAAAGAAACCAAGGTGAAACATCTTC	720
198	-----GlnGlyIysGlnGlnAspAsnAspGlyAsnAsnIlePhe	212
721	AGCGGCTTACGCGGAGTTCCTGGAAACAAGCCTTCAGGTTCCAGCTTCACGACACACAGATAGTG	780
213	SerGlyPheIysArgAspPheLeuGluAspAlaPheAsnVal---AsnArgHisIleVal	231
781	CAAAACCTAAGAGCGGACCGGAGAGTGTAAGAGAGGGAGCCATTGTGCAGACTGACGGGA	840
232	AspArgLeuGlnGlyArgAsnGluAspGluGluIysGlyAlaIleValIysValIysGly	251
841	GGCCTCAGATCTTGAGCCCGACATAGAAAG-----AGACGT	876
252	GlyLeuSerIleIleAlaProProGluArgGlnAlaArgHisGluArgGlySerArgGln	271
877	GCCGACGACGAAGACGAATACGATGAA-----GAT	906
272	GluGluAspGluAspGluIysGluGluArgGlnProSerHisIysSerArgArgAsp	291
907	GAATATGAATACGATGAAGAGGATAGA-----AGCGTGGCGAGGGAACGACGAGCAGG	960
292	GluAspGluAspAspIysGluIysArgHisSerGlnIysGlyGlnSerArgArgGlnGly	311
961	GGGAATGGTATTCAAGACGACGATCTGCACCGCAAGTGTCTTAAAGAAGACATTTGTAGNAAC	1020
312	AspAsnGlyLeuGluGluThrValCysThrAlaIysLeuArgAlaAsnIleGlySerSer	331
1021	AGATCCCTTGACATCTACAAACCTCAAGCTGGTTCCTCAAACTGCCAACGATCTCAAC	1080
332	ProSerProAspIleTyrAsnProGlnAlaGlyArgIleIysThrValThrSerLeuAsp	351
1081	CTTCTTAATACTTAGTGGCTTGGACCTAGTGTGTAATATGGAATCTCTACAGGAATGCA	1140
352	LeuProValLeuArgTrpLeuIysLeuSerAlaGluHisGlySerLeuHisIysAsnAla	371
1141	TTGTTTGTGCTCACTCAACACCAACGACGACGATCATATATTCGATTGAGGGGACGG	1200
372	MetPheValProHisTyrAsnLeuAsnAlaAsnSerValIleTyrAlaLeuIysGlyArg	391
1201	GCTCACGTGCAAGTCGTGGACACGACGCAACGAGAGTGTAACGACGAGGAGCTTCAAGAG	1260
392	AlaArgLeuGlnValValAsnCysAsnGlyAsnThrValPheAspGlyGluLeuGluAla	411
1261	GGTCACGTGCTTGTGGTCCACAAACTTCGCGCTGCTGGTGAAGTCCACGAGCAGAAC	1320
412	GlyArgAlaLeuThrValProGlnAsnTyrAlaValAlaIysSerLeuSerGluArg	431
1321	TTGCAATACGTGGCATTCAGACAGACTCAAGGCCCGACGATAGCCAACTTCGCGCGGTGAA	1380
432	PheThrTyrValAlaPheIysThrAspAspArgAlaSerIleAlaArgLeuAlaGlyThr	451

QY	1384	TCGTCATAGATAA	CTCGCGGAGGAGTGGTTGCAAAATTCATATGGCTCCAAAGGAG	1443
Db	472	SerValIleAsnAsnLeuProLeuAspValValAlaAlaThrPheAsnLeuGlnArgAsn	491	
QY	1444	CAGGCAAGGAGCTTAAAGAACAAACCCCTTCAAGTTCTTCTGTTCCACCGCTCTCAGCAG	1503	
Db	492	GluAlaArgGlnLeuLysSerAsnAsnProPheLysPheLeuValProAlaArgGluSer	511	
QY	1504	TCCTCCGAGGCTGGCT	1521	
Db	512	GluAsnArgAlaSerAla	517	
RESULT 9				
S49877				
legumin A precursor - Vicia narbonensis				
C:Species: Vicia narbonensis				
C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004				
C:Accession: S49877				
R:Nong, V.; Becker, C.; Muentz, K.				
submitted to the EMBL Data Library, November 1994				
A:Description: PCR cloning of legumin cDNA from Vicia narbonensis.				
A:Reference number: S49877				
A:Accession: S49877				
A:Molecule type: mRNA				
A:Residues: 1-482 <NON>				
A:Cross-references: UNIPROT:O41676; EMBL:Z46803; NID:g600107; PIDN:CAA86824.1; PID:g6001				
A:Experimental source: tissue type cotyledon				
C:Superfamily: glycinin				
C:Keywords: seed; storage protein				
F:1-21/Domain: signal sequence #status predicted <SIG>				
F:122-482/Product: legumin A #status predicted <MAT>				
Alignment Scores:				
Pred. No.:	1,78e-98	Length:	482	
Score:	1431.50	Matches:	284	
Percent Similarity:	68.23%	Conservative:	66	
Best Local Similarity:	55.36%	Mismatches:	104	
Query Match:	53.20%	Indels:	59	
DB:	2	Gaps:	7	
US-10-728-051-3 (1-1524) x S49877 (1-482)				
QY	1	CGGAGCAACCGGAGGAGAACGCGTTCAGGCTTCAGCGCCTCAATGCGCAGACCTTGAC	60	
Db	23	ArgGluGlnSerGlnAsnGluCysGlnLeuGluArgLeuAspAlaLeuGluProAsp	42	
QY	61	AATCGCATTTGAATCAGAGGGCGTTTACATTGAGACTTGGAAACCCCAACACCAGGAGTTC	120	
Db	43	AsnArgGluSerGluGlyLeuIleGluThrTrpAsnProAsnAsnArgGlnPhe	62	
QY	121	GAATCGCGCGGCTCGCTCTCTCGCTTAGTCTCGCCGCAACGCGCTTCGTAGGCCT	180	
Db	63	ArgCysAlaGlyValAlaLeuSerArgValThrLeuGlnArgAsnAlaLeuArgPro	82	
QY	181	TTCTACTCCAATGCTCCCGAGAGATCTTCATCCAGCAAGGAGGATATTGGGTG	240	
Db	83	TyrTyrSerAsnAlaProGlnGluIleTyrIleGlnGlnGlyAsnGlyTyrPheGlyVal	102	
QY	241	ATATTCCCTGGTTCCTAGACACTATGAGAGCCTCACACAAAGGTCGTGATCTCAG	300	
Db	103	ValPheProGlyCysProGluThrPheGluGluPro-----GlnGluSer	117	
QY	301	TCCCAAGACCAACCAAGAGCTCTCCAAGGAGAGAACCAACCAACGAGATAGT	360	
Db	118	GluGlnArgGluArgArg-----TyrArgAspSer	128	
QY	361	CACCAGAGGTGACCGTTTCGATGAGGTGATCTCATTCAGTTCCACCGGTTCCT	420	
Db	129	HisGlnLysValAsnArgPheArgGluGlyAspIleAlaValProThrGlyAsnVal	148	
QY	421	TTCTGGCTTACCAACCAACCAACCACTGATGTTGTTGTTCTTCTTACTGACACCAAC	480	

Db	149	LeuTrpMetTyrAsnAspGlnAspThrProValIleAlaIleSerLeuThrAspThrGly	168	
QY	481	AACAACGACAAACAGCTTTCATCAGTTCCCGCAGAGATTCAATTTGGCTGGGAACGAG	540	
Db	169	SerSerAsnAsnGlnLeuAspGlnIleProArgArgPheTyrLeuAlaGlyAsnGlnGlu	188	
QY	541	CAAGAGTTCTTAAGGTACCAGCAACAAACAGACAGACAAAGACGAGAGACTTACCATAT	600	
Db	189	GlnGluPheLeuArgTyrGln-----	195	
QY	601	AGCCCATACAGCCCGCAAAAGTCAGCCTAGACAAAGAGCGGTGAATTTAGCCCTCGAGGA	660	
Db	195	-----	195	
QY	661	CAGCACAGCCGACAGAGAAACGAGCAGGA---CAAGAAGAGAGAAACGAGGTGAAACATC	717	
Db	196	-----ArgGluGlnGlyGlyLysGlnGluGlnAsnAspGlyAsnAla	211	
QY	718	TTACAGCGCTTACGCGCGAGTTCTTGGACACAGCCTTCCAGGTTGACGACAGACAGATA	777	
Db	212	PheSerGlyPheLysArgAspPheLeuGluAspAlaLeuAsnVal---AsnArgHisIle	230	
QY	778	GTGCAAAACCTAAGAGCGCAGACCGAGAGTGAAGAAGAGGAGCCATTGTGCAGGTGAGG	837	
Db	231	ValAspArgLeuGlnGlyArgAsnGluAspGluGlyAlaIleValLysValLys	250	
QY	838	GGAGCGCTCAGAAATCTTGGACCCAGATAGAAAGAGAGCTGCCAGCAGAGAGGGAATAC	897	
Db	251	GlyGlyLeuSerIleIleThrProGluArgGlnArgLysSerArgGlnGluGlu---	269	
QY	898	GATGAAGTGAATATGAATACGATAGAGAGATAGAGCGGTGCGAGCGGGAAGCAGAGGC	957	
Db	270	AspGluAspGluLysGluGluArgGlnProSerArgArgArgAspGluSerGlnLysGly	289	
QY	958	AGGGGG-----AATGGTATTGAAGAGACGATCTGCACCGCAGGTGCTAAA	1002	
Db	290	GluSerArgArgHisGlyAspAsnGlyLeuGluGluThrValCysThrAlaLysLeuArg	309	
QY	1003	AAGAACTTGTAGAAACAGATCCCTGACATCTACACCTCAAGCTGTTGTTCACTCAAA	1062	
Db	310	ValAsnIleGlySerProSerProAspIleTyrAsnProGlnAlaGlyArgIleAsn	329	
QY	1063	ACTGCGCAACGATCTCAACCTTCTAATACTTAGTGGCTTGGACCTAGTGTGAATATGGA	1122	
Db	330	ThrValThrSerLeuAspLeuProValLeuArgTrpLeuLysLeuSerAlaGluHisGly	349	
QY	1123	AATCTCTACAGAAATGCTTGTTCCTCCTACTACACCAACGACGACAGCATCATA	1182	
Db	350	SerLeuArgLysAsnAlaLeuIleValProHisTyrAsnArgAsnAlaAsnSerValIle	369	
QY	1183	TATCGATTGAGGGACGGGCTCACGTCGAAGTCGTGGACAGCAGCGCAACAGAGTGTAC	1242	
Db	370	TyrAlaLeuLysGlyArgAlaArgLeuGlnValValAsnCysAsnGlyAsnThrValPhe	389	
QY	1243	GACGAGGAGCTTCAAGAGGTCACGTCGTTGTGGTCCACAGAACTTCGCGTCCCTGGA	1302	
Db	390	AspGlyGluLeuGluAlaGlyArgAlaLeuThrValProGlnAsnTyrAlaValAlaAla	409	
QY	1303	AAGTCCACGAGCGAAGCTTCGAATACGTGGCATTCACAGACAGACTCAAGGCCCGCAGCATA	1362	
Db	410	LysSerLeuSerGluArgPheThrTyrValAlaPheLysThrAsnAspArgAspGlyIle	429	
QY	1363	GCCAACTCGCGCGTGAATACTCCGTCATAGATACCTCGCGGAGAGGTGTTGCAAT	1422	
Db	430	AlaArgLeuAlaGlyThrSerSerValIleAsnAspLeuProLeuAspValValAlaAla	449	
QY	1423	TCATATGGCTTCAAAGGAGGAGCGAGCGAGCTTTAAGAACCAACCCCTTCAAGTTC	1482	
Db	450	ThrPheAsnLeuGlnArgAsnGluAlaArgGlnLeuLysSerAsnAsnProPheLysLeu	469	
QY	1483	TTCTGTTCCACCGTCTCAGCAGTCTCCGAGGCGTGTGCT	1521	
Db	470	LeuValProProArgGluSerGluLysArgAlaSerAla	482	

RESULT 10

S11003
glycinin G3 precursor - soybean
C:Species: Glycine max (soybean)
C:Date: 21-Nov-1993 #sequence_revision 19-Jan-1996 #text_change 03-May-1996
C:Accession: S11003
R:Nielsen, N.C.; Dickinson, C.D.; Cho, T.J.; Thanh, V.H.; Scallan, B.J.; Fischer, R.L.;
Plant Cell 1, 313-328, 1989
A:Title: Characterization of the glycinin gene family in soybean.
A:Reference number: S10851; MUID:92393391; PMID:2485233
A:Accession: S11003
A:Molecule type: DNA
A:Residues: 1-484 <NIE>
A:Experimental source: variety Dare
C:Genetics:
A:Gene: Gy3
C:Superfamily: glycinin
C:Keywords: storage protein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-484/Product: glycinin G3 #status predicted <MAT>

Alignment Scores:
Pred. No.: 3,25e-98 Length: 484
Score: 1428.00 Matches: 290
Percent Similarity: 70.33% Conservative: 68
Best Local Similarity: 56.97% Mismatches: 95
Query Match: 53.07% Indels: 56
DB: 2 Gaps: 9

US-10-728-051-3 (1-1524) x S11003 (1-484)

Qy	1	CGGCAGCAACCGGAGGAAACGGCTCCAGTTCAGCGCCCTCAATGCGCAGACCTGAC	60
Db	23	ArgGluGlnProGInGInAsnGlnCuysGlnIleGlnArgLeuAsnAlaLeuIlysProAsp	42
Qy	61	AATCGCATTGAATCAGAGGGCGGTTACATTCAGACTTCGAACCCCAACCAACAGGAGTTC	120
Db	43	AsnArgIleGluSerGluGluGlyPheIleGluThrTrpAsnProAsnIlysProPhe	62
Qy	121	GAATGCGCGCGCTCGCCCTCTCTCGTGTAGTCTCTCCGCCGCAACGCCCTTCGTAGGCCT	180
Db	63	GlnCySAlaGlyValAlaLeuSerArgCySThrLeuAsnArgAsnAlaLeuArgArgpro	82
Qy	181	TTCTACTCCATGCTCCCGAGGATCTTCATCCAGCAAGGAGGGGATCTTTGGGTG	240
Db	83	SerTyrThrAsnAlaProGlnGluIleTyrIleGlnGlnGlySerGlyIlePheGlyMet	102
Qy	241	ATATTCCCTGGTGTCTCTAGACACTATCAAGAGCCTCACACAGAGTCTCGATCTCAG	300
Db	103	IlePheProGlyCySProSerThrPheGluGluProGlnGlnIlysGly-----Gln	119
Qy	301	TCCCAAGACCAACCAAGACGTCTCCACGAGAGAACCACCAAGCAACAGACGATAGT	360
Db	120	SerSerArgPro-----GlnAspArg 126	
Qy	361	CACCAAGAGGTGCACCGTTTCGATGAGGTGATCTCATTCAGTTCGCCACCGGTCTGCT	420
Db	127	HisGlnIlySerIleTyrHisPheArgGluGlyAspLeuIleAlaValProThrGlyPheAla	146
Qy	421	TTCTGGGTCTCAACAGCACACGACACTGATGTGTGTGCTTTCTCTTACTGACACCAAC	480
Db	147	TyrTrpMetTyrAsnAsnGluAspThrProValValAlaValSerLeuIleaspThrAsn	166
Qy	481	AACAAACGACCAACCGCTTGATCAGTTCGCCAGGAGATTCATTTGGCTGGGAACACGGAG	540
Db	167	SerPheGlnAsnGlnLeuAspGlnMetProArgArgPheTyrLeuAlaGlyAsnGlnGlu	186
Qy	541	CAAGAGTTCTTAAGTACCAGCAACAAAGCAGACAAAGCAGACGAAGAAGCTTACCATAT	600
Db	187	GlnGluPheLeuGlnTyrGlnProGlnIlySglnGlnGly-----	199
Qy	601	AGCCCATACAGCCCGCAAGTCAGGCTAGACAAGAAGAGCGTGAATTTAGCCCTCGAGA	660

Db 200 -----Gly 200

Qy 661 CAGCACAGCGCAGAGAACGAGCAGGACAAGAAAGAAAACGAAGGTGGAAACATCTTC 720

Db 201 ThrGlnSerGlnLysGlyAsArgGInGluGluGluAenGluGlySerIleLeu 220

Qy 721 AGCGGCTTACGCCGAGATTCTTGGAACAAGCCCTTCAGGTTGCACACAGACAGATAGTG 780

Db 221 SerGlyPheAlaProGluPheLeuGluHisAlaPheValVal---AspArgGlnIleVal 239

Qy 781 CAAACTTAAGCGCGAGACCAGAGCTGAAGAAGAGGAGCGCATCTGCAGAGTAGGGGA 840

Db 240 ArgLysLeuGlnGlyLysGluGluGluGluGlyAlaIleValThrValLysGly 259

Qy 841 GGCTCAGAAATCTTGAGCCCA-----GATGAAAAGACGTCGCCGACGAAGAAGAG 891

Db 260 GlyLeuSerValIleSerProThrGluGluGlnGlnArgProGluGluGlu 279

Qy 892 GAATACGATCAAGATGAATATGAATACGATGAGAGGATAGAGCGCTGCGCGGGNAGC 951

Db 280 LysProAsp-----CysAspGluLysAspLysHisCysGlnSerGlnSer 294

Qy 952 AGGAGCGAGGGGAATGGTATTGAAGAGACGATCTGCACCGCAAGTGTAAAAAGAACATT 1011

Db 295 Arg-----AsnGlyIleAspGluThrIleCysThrMetArgLeuArgHisAenIle 311

Qy 1012 GGTAGAAAACAGATCCCTCGACATCTCAACCCCTCAAGCTGGTTCACTCAAACCTGCCAAC 1071

Db 312 GlyGlnThr-SerSerProAspIlePheAsnProGlnAlaGlySerIleThrThrAlaThr 331

Qy 1072 GATCTCAACCTTCTAATACTTAGTGGCTGGACCTAGTCTGAATATGGAATCTCTAC 1131

Db 332 SerLeuAspPheProAlaLeuSerTrpLeuLysLeuSerAlaGlnPheGlySerLeuArg 351

Qy 1132 AGGAATGCATTGTTTGTGCTCACTCAACACCAACGCAACAGCATCATATATCGATTG 1191

Db 352 LysAsnAlaMetPheValProHisIleTyrAsnLeuAsnAlaAsnSerIleIleTyrAlaLeu 371

Qy 1192 AGGGAGCGGCTCAGCTGCAGTCTGGACAGCAACGCGCAACAGAGTGTACGACGAGGAG 1251

Db 372 AsnGlyArgAlaLeuValGlnValValAsnAsnGlyGluArgValPheAspGlyGlu 391

Qy 1252 CTTCAAGAGGGTCACGTGTTGGTGCCACAGAACTTCGCCGTGCTGGAAGTCCCAG 1311

Db 392 LeuGlnGlyGlnValLeuIleValProGlnAsnPheAlaValAlaalaArgSerGln 411

Qy 1312 AGCGAGAACTTCGAATACGTGGCATTCACAGACAGACTCAAGGCCCGACATAGCCAACCTC 1371

Db 412 SerAspAsnPheGluTyrValSerPheLysThrAsnAspArgProSerIleGlyAenLeu 431

Qy 1372 CGCGGTGAAACTCCGTCATAGATAACCTCGCGGAGGAGGTGTTGCMAATTCATATGGC 1431

Db 432 AlaGlyAlaAsnSerLeuLeuAsnAlaLeuProGluGluValIleGlnGlnThrPheAsn 451

Qy 1432 CTCCAAAGGGAGCAGCAAGGCAGCTTAAGAACAAACACCCCTTCAAGTTCCTCGTT--- 1488

Db 452 LeuArgGlnSerGlnValSerGluLeuLysTyrGluGlyAsnTrpGlyProLeuValAsn 471

Qy 1489 CCACCGTCTCAGCAG---TCTCCGAGG 1512

Db 472 ProGluSerGlnGlnGlySerProArg 480

RESULT 11

S14393

legumin A2 precursor - fava bean

C;Species: Vicia faba (fava bean)

C;Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004

R;Accession: S14393

R;Schlesier, B.; Basuener, R.; van Hai, N.; Muentz, K.

Nucleic Acids Res. 18, 7146, 1990

A;Title: The cDNA derived primary structure of two distinct legumin A subunit p

A;Reference number: S14392; PMID:91088307; PMID:2263491

Mon Apr 11 05:25:05 2005

A;Accession: S14393
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-500 <SCH>
A;Cross-references: UNIPROT:Q99304; EMBL:X55014; NID:g22007; PIDN:CAA38758.1; PID:g22008
C;Superfamily: glycinin

Alignment Scores:
Pred. No.: 3,29e-97 Length: 500
Score: 1414.50 Matches: 280
Percent Similarity: 67.30% Conservative: 76
Best Local Similarity: 52.93% Mismatches: 100
Query Match: 52.56% Indels: 73
DB: 2 Gaps: 8

US-10-728-051-3 (1-1524) x S14393 (1-500)

Qy 1 CGGCAGCAACCGGAGGAGACGGTCCAGTTCACGCGCTCAATCGCGAGAGACTGCAC 60
Db 23 ArgGluGlnSerGlnGlnAenGluCysGlnLeuGluArgLeuAspAlaLeuGluProAsp 42
Qy 61 AATCGCATTTGAATCAGAGGCGGTTCATATTGAGACTTGGAAACCCCAACACAGGAGTTC 120
Db 43 AsnArgileGluSerGluGlyGlyLeuileGluThrTriAsnProAsnAsnArgGlnPhe 62
Qy 121 GAATCGCGCGCGTCCGCTCTCTCGCTTAGTCTCGCGCGCAACGCCCTTCGTAGGCT 180
Db 63 ArgCysAlaSerValAlaLeuSerArgAlaThrLeuGlnArgAsnAlaLeuArgArgPro 82
Qy 181 TTCTACTCCAAATGCTCCCGAGGAGATCTTCATCAGCAAGAAAGGGGATATTTGGGTG 240
Db 83 TyrTyrSerAsnAlaProGlnGluileTyrileGlnGlyAsnGlyTyrPheGlyMet 102
Qy 241 ATATTCCTGTTTCTTAGACACTATGAGAGGCTCAGACACAGAGTCTCGATCTCAG 300
Db 103 ValPheProSerCysProGluThrPheGluGluPro-----GlnGln 116
Qy 301 TCCCAAGACACCAAGAGCGTCTCCAAGGAGAGACCAAGCAACAGCAAGAGATAGT 360
Db 117 SerGlu-----GlnGlyGlu-----GlyGlyArgTyrArgAspSer 128
Qy 361 CACCAAGAGTGCACCGTTTCGATGAGGTGATCTCATTTGAGTTCACCGGTGTGCT 420
Db 129 HisGlnLysValAsnArgPheArgGluGlyAspilleAlaValProThrGlyileVal 148
Qy 421 TTCTGCTCTCAACGACACGACACTGATGTTGTTGTTCTTCTTCTTACTGACACCAAC 480
Db 149 PheTrpMetTyrAsnAspGlnAspThrProValileAlaileSerLeuThrAspilleGly 168
Qy 481 AACCAACGACCAACCGCTTATCAGTTCCTCCAGGAGATTCATTTGGCTGGGAACACCGAG 540
Db 169 SerSerAsnAsnGlnLeuAspGlnMetProArgArgPheTyrLeuAlaGlyAsnGlnGlu 188
Qy 541 CAAGAGTTCCTTAAGTACAGCAACAAAGCAGACAAAGCAGACAAAGAGCTTACCATAT 600
Db 189 GlnGluPheLeuArgTyrGlnHisGln-----GlnGln----- 197
Qy 601 AGCCCATACAGCCCGCAAGTCCAGCTAGACAAAGAGAGCGTGAATTTAGCCCTCGAGGA 660
Db 197 ----- 197
Qy 661 CAGCACAGCGGAGAAACGACAGGACNAGAAAGAAAGAAAGAGTGGAAACATCTTC 720
Db 198 -----GlnGlyGlyLysGluGluGlnAspAsnAspGlyAsnAsnillePhe 212
Qy 721 AGCGGCTTCAGCGCGAGTTCCTGGAAACAAAGCGCTTCCAGTTTGACGACAGACAGATGTG 780
Db 213 SerGlyPheLysArgAspPheLeuGluAspAlaLeuAsnVal---AsnArgHisileVal 231
Qy 781 CAAACCTAAGAGCGGAGACCGAGAGTGAAGAGAGGAGCGCATTTGTGACGTAGAGGGA 840
Db 232 AspArgLeuGlnGlyArgAsnGlnAspGluGlyAlaileValLysValLysGly 251

Qy 841 GGCCTCAGATCTTGAGCCAGATAGAAAG-----AGCGTCCGACGAA 885
Db 252 GlyLeuSerileThrProGluArgGlnAlaArgHisProArgGlySerArgGln 271
Qy 886 GAAGAGAAATACATCAAGATGATATGAA----- 915
Db 272 GluGluAspGluAspGluAspGluGluGluArgGlnProSerHisLysSerArg 291
Qy 916 -----TACGATCAAGAGGATAGA-----AGCGTGGCAGGGGAAACAGA 954
Db 292 ArgGlyGluAspGluAspAspLysGluLysArgHisSerGlnLysGlyGluSerArgArg 311
Qy 955 GCGAGGGGAAATGATTTGAAGAGAGATCTGACCGCAAGTCTAAAGAACATCTGCT 1014
Db 312 HisGlyAspAsnGlyLeuGluGluThrValCysThrAlaLysLeuArgLeuAsnilleGly 331
Qy 1015 AGAAACAGATCCCTGACATCTACACCTCAAGCTGGTGTCTCACTCAAACTGCCAACGAT 1074
Db 332 SerSerSerSerProAspIleTyrAsnProGlnAlaGlyArgileLysThrValThrSer 351
Qy 1075 CTCAACCTTCTTAATCTTAGTGTGCTTGGACCTTAGTGTCTGAATATGAAATCTCTACAGG 1134
Db 352 LeuAspLeuProValLeuArgTyrLeuLysLeuSerAlaGluHisGlySerLeuArgLys 371
Qy 1135 AATGATTTGTTGCTGCTCACTACACCAACGACACACAGCATCATATATCGATTGAGG 1194
Db 372 AsnAlaMetPheValProHisTyrAsnLeuAsnAlaAsnSerileLeuTyrAlaLeuLys 391
Qy 1195 GCGCGGCTCAGCTGCAAGTCTCGGACAGCAACCGCAACGAGTGTACGACGAGGAGCTT 1254
Db 392 GlyArgAlaArgLeuGlnValValAsnCysAsnGlyAsnThrValPheAspGluGluLeu 411
Qy 1255 CAAGAGGTTCAGTGTGCTGCTGCCACAGACTTCCCGCTCGCTGGAAGTCCACAGC 1314
Db 412 GluAlaGlyArgAlaLeuThrValProGlnAsnTyrAlaValAlaLysSerLeuSer 431
Qy 1315 GAGAACTTCGAATCGTGCATTCAGACAGACTCAAGGCCACGATAGCCACCTCGCC 1374
Db 432 AspArgPheThrTyrValAlaPheLysThrAsnAspArgAlaGlyileAlaArgLeuAla 451
Qy 1375 GGTGAAAACTCGTCATAGATAACCTCGCGAGAGGTTGTGCAATTCATATGGCTC 1434
Db 452 GlyThrSerSerValileAsnAspMetProValAlaValAlaThrPheAsnLeu 471
Qy 1435 CAAAGGAGCGAGGAGGAGCGCTTAAGAAACAAACACCCCTTCAAGTCTTCGTTCCACCG 1494
Db 472 GluArgAsnGluAlaArgGlnLeuLysSerAsnAsnProPheLysPheLeuValProPro 491
Qy 1495 TCTCAGCAGTCTCCGAGGCGTGTGCT 1521
Db 492 ArgGluSerGlnLysArgAlaSerAla 500

RESULT 12

T06452
probable legumin A precursor - garden pea
C;Species: Pisum sativum [garden pea]
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 21-Jul-2000
C;Accession: T06452
R;Lycett, G.W.; Croy, R.D.; Shirsat, A.H.; Richards, D.M.; Boulter, D.
Nucleic Acids Res. 13, 6733-6743, 1985
A;Title: The 5'-flanking regions of three legumin genes: comparison of the DNA sequences
A;Reference number: Z15687; MUID:86041868; PMID:2997721
A;Accession: T06452
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-507 <LYC>
A;Cross-references: EMBL:X02982; NID:g20777; PIDN:CAA26720.1; PID:g4379378
C;Genetics:
A;Introns: 96/1; 179/3; 388/3
C;Superfamily: glycinin
C;Keywords: seed; storage protein
P;1-21/Domain: signal sequence #status predicted <SIG>
P;22-332/Product: legumin A, alpha chain #status predicted <ALP>

DB:	2	Gaps:	11
DB:	US-10-728-051-3 (1-1524) x S14392 (1-497)		
QY	25	TGCCAGTTCACGGCTCAATCGCGAGACCTGACAAATCGCATTGAATCAGAGGGCGGT	84
Db	1	CysGlnLeuGluArgLeuAspAlaLeuGluProAspAsnArgIleGluSerGluGlyGly	20
QY	85	TACATTGAGACTTGGAAACCCCAACACAGGAGTTCGAATGCGCGCGGTGCTCTCTCT	144
Db	21	LeuIleGluThrTrpAsnProAsnAsnArgGlnPheArgCysAlaArgValAlaLeuSer	40
QY	145	CGTTAGTCTCCGCGCAACGCCCTTCGTAGGCTTCTACTCCAAATGCTCCCGCAGAG	204
Db	41	ArgAlaThrLeuGlnArgAsnAlaLeuArgA-gProTyrTyrSerAsnAlaProGlnGlu	60
QY	205	ATCTTCATCCAGCAAGGAGGATACTTTGGTGTGATATTCCTCGTGTCTCTAGACAC	264
Db	61	IleTyrIleGlnGlnGlyAsnGlyTyrPheGlyMetValPheProSerCysProGluThr	80
QY	265	TATGAAGAGCTCACACAAAGTCTCGATCTCAGTCCCAAGACCCCAAGAGCTCTC	324
Db	81	PheGluGluPro-----GlnGlnSerGlu-----	88
QY	325	CAAGGAGAACACCAAGCCCAACAGACAGATAGTACACAGAAAGTGCGCGTTCGAT	384
Db	89	GlnGlyGlu-----GlyArgArgTyrArgAspSerHisGlnLysValAsnArgPheArg	106
QY	385	GAGGGTGATCTCATTTGAGTTCCTCCCGGTGTGCTTCTGCTCTACAAACGACACGAC	444
Db	107	GlnGlyAspIleIleAlaValProThrGlyIleValPheTrpMetTyrAsnAspGlnAsp	126
QY	445	ACTGATGTGTTGCTGTTCTCTTACTGACCAACCAACAGACAGCTTGTATCAG	504
Db	127	IleProValIleAlaIleSerLeuThrAspThrGlySerSerAsnAsnGlnLeuAspGln	146
QY	505	TTCCCGAGGAGATCAATTTGGCTGGGAACACGAGCAAGAGTCTTAAGGTACACGAA	564
Db	147	MetProArgArgPheTyrLeuAlaGlyAsnGlnGluGlnPheLeuArgTyrGlnHis	166
QY	565	CAAGACGACAAACGACGAGAGAGTTCACCATATAGCCATACAGCCCGCAAGTCAG	624
Db	167	GlnGlnGlyValLysGluGlu-----GlnAspAsn	176
QY	625	CCTAGACAAAGAGCGTGAATTTAGCCCTCGAGGACGACAGCCCGCAGACGACGCA	684
Db	177	AspGlyAsnGlnGluGlnGluPheLeu--ArgTyrGlnHis-----ArgGlnGlyVal	193
QY	685	GGACAAGAGAGAAACAAAGGTGGAAACATCTTCAGCGGCTTCACGCGGAGTTCCTG	744
Db	194	LysGluGluGlnAspAsnAspGlyAsnAsnIlePheSerGlyPheAsnArgAspPheLeu	213
QY	745	GAACAAGCTCCAGGTTGACGACAGACAGATAGTGCAAACTTAAGAGGCGAGCCGAG	804
Db	214	GluAspAlaPheAsnVal---AsnArgHisIleValAspArgLeuGlnGly-----Arg	230
QY	805	AGTGAAGAGAGGGAGCCATTGTGACAGTGGAGGGAGCCCTCAGAACTTGAGCCCGAT	864
Db	231	AsnGluGluArgGlyAlaIleValLysValLysGlyGlyLeuSerIleIleThrProPro	250
QY	865	AGAAAGAGA-----CGTCCGCAAGAGAGAGGAATACGATGAA	903
Db	251	GluArgGlnAlaArgHisProArgGlySerArgGlnGluGluAspGluAspGlu	270
QY	904	GATGAATATGAA-----TACGATGAAGAG	927
Db	271	AspGlyLysGluGluArgGlnProSerHisHisLysSerArgArgGlyGluAspGluAsp	290
QY	928	GATAGA-----AGGCGTGGCAGGGAACGACAGAGGCGAGGGGGAATGTATT	972
Db	291	AspLysGluLysArgHisSerGlnLysGlyGlyLeuSerArgArgHisGlyAspAsnGlyLeu	310
QY	973	GAAGACACGATCTGCACCGCAAGTGTCTAAAGAAACATTTGGTGAAGAACAGATCCCGCTGAC	1032

Db	311	GluGluThrValCysThrAlaLysLeuArgLeuAsnIleGlySerSerSerProAsp	330
QY	1033	ATCTACACCTCAAGCTGGTTCACCTCAAACTGCCAACGATCTCAACCTTCTTAATACT	1092
Db	331	IleTyrAsnProGlnAlaGlyArgIleLysThrValThrSerValAspLeuProValLeu	350
QY	1093	AGTGGCTTGGACCTAGTGTGAATATGGAATCTCTACAGGAATGATGTTTGTCTGCT	1152
Db	351	ArgTrpLeuLysLeuSerAlaGluHisGlySerLeuArgLysAsnAlaMetPheValPro	370
QY	1153	CATTACAAACACCGCACACACATCATATATCGATTGAGGGGAGCGGCTCACGTGCAA	1212
Db	371	HisTyrAsnLeuAsnAlaAsnSerValLeuTyrAlaLeuLysGlyArgAlaArgLeuGln	390
QY	1213	GTGCTGCACAGCAACGCGACAGAGTGTACGACGAGGAGCTTCAAGAGGGTCCAGTCTT	1272
Db	391	ValValAsnCysAsnGlyAsnThrValPheAspGlyGluLeuGluAlaGlyArgAlaLeu	410
QY	1273	GTGCTGCACAGCAACCTTCGCGTGGTGGAAAGTCCAGAGCGAGAACTTCGATACGTG	1332
Db	411	ThrValProGlnAsnTyrValValAlaAlaLysSerLeuSerAspArgPheThrTyrVal	430
QY	1333	GCATTGACAGACAGACTCAAGCCGACATAGCCAACTCCCGGTGAAAACTCCGTCATA	1392
Db	431	AlaPheLysThrAsnAspArgAlaGlyIleAlaArgLeuAlaGlyThrSerSerValIle	450
QY	1393	GATAACCTCGCGGAGGAGTGGTTCGCAATTCATATATGCGCTCCAAAGGAGCGGCAAGG	1452
Db	451	AsnAspLeuProLeuAspValValAlaAlaThrPheAsnLeuGluArgAsnGluAlaArg	470
QY	1453	CAGCTTAAGAACACCAACCCCTTCAAGTCTTCGTTCCACCGTCTCAG	1500
Db	471	GlnLeuLysPheAsnAsnProSerArgPheLeuValProArgGlu	486
RESULT 14			
S51941			
prunin 1 precursor - almond			
N:Alternate names: salt-soluble globulin; seed storage protein			
C:Species: Prunus dulcis (almond)			
C:Date: 28-Oct-1996 #sequence_revision 07-Feb-1997 #text_change 09-Jul-2004			
C:Accession: S51941; S42474			
R:Garcia-Mas, J.; Messegue, R.; Arus, P.; Puigdomenech, P.			
Plant Mol. Biol. 27, 205-210, 1995			
A:Title: Molecular characterization of cDNAs corresponding to genes expressed during alc			
A:Reference number: S51940; MUID:95170003; PMID:7865791			
A:Accession: S51941			
A>Status: nucleic acid sequence not shown			
A:Molecule type: mRNA			
A:Residues: 1-551 <GAR>			
A:Cross-references: UNIPROT:Q43607; EMBL:X78119			
R:Garcia-Mas, J.; Messegue, R.; Arus, P.; Puigdomenech, P.			
submitted to the EMBL Data Library, March 1994			
A:Description: Molecular characterization of cDNAs corresponding to proteins related to			
A:Reference number: S42473			
A:Accession: S42474			
A:Molecule type: mRNA			
A:Residues: 1-60, G' 62-551 <GAW>			
A:Cross-references: EMBL:X78119; NID:g460805; PIDN:CAA55009.1; PID:g460806			
A>Note: the source is designated as Prunus amygdalus			
C:Superfamily: glycinin			
C:Keywords: seed; storage protein			
F:1-20/Domain: signal sequence #status predicted <SIG>			
F:21-367/Product: prunin 1 alpha chain #status predicted <MAT1>			
F:368-551/Product: prunin 1 beta chain #status predicted <MAT2>			
F:108-374/Disulfide bonds: #status predicted			
Alignment Scores:			
Pred. No.:	7,47e-72	Length:	551
Score:	1073.50	Matches:	228
Percent Similarity:	57.22%	Conservative:	89
Best Local Similarity:	41.16%	Mismatches:	160

Query Match:	39.89%	Indels:	77
DB:	2	Gaps:	11
US-10-728-051-3 (1-1524) x S51941 (1-551)			
QY	4	CAGCAACGGAGGAGCGGTCCAGCTTCAGCGCCTCAATGCGCAGAGACCTGCAAT	63
DB	25	GlnLeuSerProGlnAsnGlnCysGlnLeuAsnGlnLeuGlnAlaArgGluProAspAsn	44
QY	64	CGCATTAATCAGAGGCGGTTCATTGAGACTTGGAAACCCCAACACAGGAGTTCGAA	123
DB	45	ArgileGlnAlaGluAlaGlyGlnIleGluThrTrpAsnPheAsnGlnGluAspPheGln	64
QY	124	TGGCGCGGCTCGCCTCTCTCGTCTAGTCTCGCGCGCAACGCCCTTCGTAGCCCTTC	183
DB	65	CysAlaGlyValAlaAlaSerArgIleThrIleGlnArgAsnGlyLeuHisLeuProSer	84
QY	184	TACTCCAATGCTCCCGAGGAGATCTTCATCCAGCAAGGAAGGGATATCTTGGTGTATA	243
DB	85	TySerAsnAlaProGlnLeuIleIleValGlnArgGlyValLeuGlyAlaVal	104
QY	244	TTCCCTGGTGTCTAGACACTATGAAGAG	273
DB	105	PheSerGlyCysProGluThrPheGluGluSerGlnGlnSerSerGlnGlnArgGln	124
QY	274	---CCTCACACACAGGTCTGATCTCAGTCCCAAGACCAACAGACGTCTCCAGGA	330
DB	125	GlnGluGlnGluGlnGluArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln	144
QY	331	GAAGACCAAGCCCAACAGCAACGA	354
DB	145	GlnGlnGlnGlnGlnGlnGluArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln	164
QY	354	---	354
DB	165	GlyArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnPhe	184
QY	355	---GATAGTACACAGAGGTCCACCGTTTCGATCAGGTGATCTCATTCGAGTT	405
DB	185	ArgGlnLeuAspArgHisGlnLysThrArgArgIleArgGluGlyAspValAlaIle	204
QY	406	CCACCGGTGTCTCTCGCTCTACACGACACGACACTGATGTGTGTGTCTCT	465
DB	205	ProAlaGlyValAlaIleTrpSerTrpAsnAspGlyAspGlnGlnLeuValAlaValAsn	224
QY	466	CTTACTGACCAACAAACAGCAACGAGCTTCAGTTCCTCCAGGAGATTCATTTG	525
DB	225	LeuPheHisValSerSerAspHisAsnGlnLeuAspGlnAsnProArgLysPheTyrLeu	244
QY	526	GCTGGGAACACGGAGAGAGTCTTAAAGTACAGCAACAAAGCAGACAAAGCAGCA	585
DB	245	AlaGlyAsnProGluAsnGluPhe	256
QY	586	AGAAGCTTACCATATACCCCATACAGCCCGCAAGTCCAGCTACAGCAAGAGCGTGAA	645
DB	257	---	266
QY	646	TTTAGCCCTCGAGCAGCAGCAGCCGC	696
DB	267	GlyArgPro---GlyGlnHisGlnGlnProPheGlyArgProArgGln---GlnGlnGln	284
QY	697	GAAGCAAGGTGGAAACATCTTCAGCGGTTCACCGCGGATTCCTGGAAACAGCGCTTC	756
DB	285	GlnGlyAsnGlyAsnAsnValPheSerGlyPheAsnThrGlnLeuLeuAlaGlnAlaLeu	304
QY	757	CAGTTCACGACACAGATAGTCCAAACCTAAGACCGCAGACCGAGAGTGAAGAGAG	816
DB	305	AsnValAsnGlu---GluThrAlaArgAsnLeuGlnGlnGlnAsnAsnArgAsnGln	323
QY	817	GGAGCCATTGTGACAGTGGAGGAGCGCTCAGAAATCTTGAAGCCAGATAGAAAGACGT	876
DB	324	-----IleIleGlnValArgGlyAsnLeuAspPheValGlnProProArgGlyArgGln	341
QY	877	GCCGCAAGAGAGAGGAATACGATGAAGATGAATATGAATACGATGAAGAGGATAGAGG	936
DB	342	GluArgGluHisGluGluArgGlnGlnGluGlnGlnGlnGlnGlnGlnGlnGlnGln	357
QY	937	CGTGGCAGGGAAGCAGAGCAGCGGGGAATGGTATTGAAGAGACGATCTCACCGCAAGT	996
DB	358	GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln	377
QY	997	GCTAAAAAGAACATTCGTAGAAACAGATCCCTGACATCTTACACCTCAAGCTGGTTCA	1056
DB	378	LeuLysGluAsnIleGlyAsnProGluArgAlaAspIlePheSerProArgAlaGlyArg	397
QY	1057	CTCAAAACTGCCACCATCTCAACCTCTTAATATTAGTGGCTTGGACCTAGTCTGAA	1116
DB	398	IleSerThrLeuAsnSerHisAsnLeuProIleLeuArgPheLeuArgLeuSerAlaGlu	417
QY	1117	TATGGAAATCTCTACAGGAATGATTTGTTGCTGCTACCTACCAACACCAACGACACAGC	1176
DB	418	ArgGlyPhePheTyrArgAsnGlyIleTyrSerProHisTrpAsnValAlaHisSer	437
QY	1177	ATCATATATGATTTGAGGGACGGGCTCAGTCAAGTCTGTCAGCAGCAACGCAACAGA	1236
DB	438	ValValTyrValIleArgGlyAsnAlaArgValGlnValValAsnGluAsnGlyAspAla	457
QY	1237	GTCAGCAGCAGGAGCTTCAAGAGGTACAGTCTGTCGTCGTCGTCGTCGTCGTCGTC	1296
DB	458	IleLeuAspGlnGluValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln	477
QY	1297	GCTGGAAGTCCACAGAGCAGAACTTCGAATACGTGCTTCAAGACAGACTCAAGGCC	1356
DB	478	IleGlnGlnAlaGlyAsnGlnGlyPheGluTyrPheAlaPheLysThrGluGluAsnAla	497
QY	1357	AGCATAGCCAACTCCCGGTGAAACTCCGTCTAGATAGTAACCTCGGAGGAGGTGGTT	1416
DB	498	PheIleAsnThrLeuAlaGlyArgThrSerPheLeuArgAlaLeuProAspGluValLeu	517
QY	1417	GCNAATTATATGCGCTCCAAAGGAGCAGGAGCAGGAGCTTAAAGCAACACACCTTC	1476
DB	518	AlaAsnAlaTyrGlnIleSerArgGluGlnAlaArgGlnLeuLysTyrAsnArgGlnGlu	537
QY	1477	AGTTCTTCTGTTCCACCGTCTCAGCAGTCTCCGAGGCGCTGTG	1518
DB	538	ThrIleAlaLeuSerSerSerGlnGlnArg---ArgAlaVal	550
RESULT 15			
S11004			
glycinin G4 precursor - soybean			
C:Species: Glycine max (soybean)			
C>Date: 21-Nov-1993 #sequence_revision 19-Jan-1996 #text_change 09-Jul-2004			
C:Accession: S11004			
R:Nielson, N.C.; Dickinson, C.D.; Cho, T.J.; Thanh, V.H.; Scallan, B.J.; Fischer, R.L.; Plant Cell 1, 313-328, 1989			
A:Title: Characterization of the glycinin gene family in soybean.			
A:Reference number: S10851; MUID:92393391; PMID:2485233			
A:Accession: S11004			
A:Molecule type: DNA			
A:Residues: 1-560 <NIE>			
A:Cross-references: UNIPROT:Q9S9D0			
A:Experimental source: variety Dare			
C:Genetics:			
A:Gene: GY4			
C:Superfamily: glycinin			
C:Keywords: storage protein			
F:1-23/Domain: signal sequence #status predicted <SIG>			
F:24-560/Product: glycinin G4 #status predicted <MAT>			
Alignment Scores:			
Pred. No.:	4,96e-70	Length:	560
Score:	1049.00	Matches:	227
Percent Similarity:	53.33%	Conservative:	77
Best Local Similarity:	39.82%	Mismatches:	158
Query Match:	38.98%	Indels:	108
DB:	2	Gaps:	10

334 AspGlnAspGlnAspGluAspGluAspGluAspGlnProArgLysSerArgGlu 353
958 -----AGGGGG 963
354 TrpArgSerLysLysThrGlnProArgArgProArgGlnGluGluProArgGluArgGly 373
964 -----AATGGTATTGAAGACAGATCTGCACCGCAAGTGTCTTAAAGAACATTT 1011
374 CysGluThrArgAsnGlyValGluGluAsnIleCysThrLeuLysLeuHisGluAsnIle 393
1012 GGTAGAAACAGATCCCTGCATCTACCAACCTCAAGCTGGTTCCTCACTCAAACTGCCAAC 1071
394 AlaArgProSerArgAlaAspPheTyrAsnProLysPheAlaClyArgIleSerThrLeuAsn 413
1072 GATCTCAACCTTCTAATCTTAGTGGCTTGGACCTAGTGTGCTGAATATGGAATCTCTAC 1131
414 SerLeuThrLeuProAlaLeuArgGlnPheGlnLeuSerAlaGlnTyrValValLeuTyr 433
1132 AGGAATCATTTGTGTCTCTACTCAACACCAACGACACGACATCATATATCGATTG 1191
434 LysAsnGlyIleTyrSerProHisTrpAsnLeuAsnAlaAsnSerValIleTyrValThr 453
1192 AGGGACGGCTCACGTGCAAGTCTGCAGCAACGCAACAGAGTGTACGACGAGGAG 1251
454 ArgGlyGlnGlyLysValArgValValAsnCysGlnGlyAsnAlaValPheAspGlyGlu 473
1252 CTTCAAGAGGGTCACTGTGTGTGTCCTCCACAGAACTTCGCGCTGGTGAAGTCCAG 1311
474 LeuArgArgGlyGlnLeuLeuValValProGlnAsnPheValValAlaGluGlnAlaGly 493
1312 ACCGAGAACTTCGATACGTGGCATTCAGACAGACTCAAGGCCACATAGCCCAACCTC 1371
494 GluGlnGlyPheGluTyrIleValPheLysThrHisAsnAlaValThrSerTyrLeu 513
1372 GCGGTGAAACTCCGTATAGATAAAGTTCGCGGAGGAGGTGGTTCGAAATTCATATGGC 1431
514 -----LysAspValPheArgAlaIleProSerGluValLeuAlaHisSerTyrAsn 530
1432 CTCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1491
531 LeuArgGlnGlnAlaArgGlnValLysAsnAsnAsnProPheSerPheLeuValPro 550
1492 CCGTCTCAGCAGTCTCCGAGGGCTGTGGCT 1521
551 ProLysGluSerGlnArgArgValValAla 560

Search completed: April 8, 2005, 05:30:59
Job time: 90.7154 secs

US-10-728-051-3 (1-1524) x S11004 (1-560)
QY 19 AAGCGCTGCAGTCTCCAGCGCTCAATGCGCAGACAGCTGACAAATCGCATTAATCAGAG 78
Db 30 AsnGluCysGlnLeuAsnAsnLeuAlaLeuGluProAspHisArgValGluSerGlu 49
QY 79 GCGGTTTACATTGAGACTTGGAAACCCCAACACAGGAGTTCGAATGCGCGCGCTGCC 138
Db 50 GlyGlyLeuIleGlnThrTrpAsnSerGlnHisProGluLeuLysCysAlaGlyValThr 69
QY 139 CTCTCTCGTCTAGTCTCCGCGCAGCGCTTCGTAGGCTTCTTACTCCAAATGCTCCC 198
Db 70 ValSerLysLeuThrLeuAsnArgAsnGlyLeuHisLeuProSerTyrSerProTyrPro 89
QY 199 CAGGAGATCTTCATCCAGCAAGAAAGGATGATCTTTGGTGTGATATTCCTGGTGTCT 258
Db 90 ArgMetIleIleIleAlaGlnGlyLysGlyAlaLeuGlyValAlaIleProGlyCysPro 109
QY 259 AGACATATGAAGAGCTTCACACAGAGTCTCGATCTCAGTCCCAAGACCAACCAAGA 318
Db 110 GluThrPheGluGluProGlnGlnSerAsnArgArgGlySerArg 125
QY 319 CGTCTCCAGGAGAGCAACCAACGACAGACAGATAGTCCACAGAAAGGTGCACCGT 378
Db 126 -----SerGlnLysGlnGlnLeuGlnAspSerHisGlnLysIleArgHis 140
QY 379 TTCGATGAGGTGATCTCATTGAGTTCACCGCGTGTGCTTCTTGGCTCTACACGAC 438
Db 141 PheAsnGluGlyAspValLeuValIleProGlyValProTyrTrpThrTyrAsnThr 160
QY 439 CACGACTGATGTGTGCTGCTTCTCTTACTGACACCAACCAACAGAGTTCCTTAAGGTAC 498
Db 161 GlyAspGluProValValAlaIleSerLeuLeuAspHisSerAsnAsnGlnLeu 180
QY 499 GATCAGTCTCCCGAGGATTCATTTGGCTGGGAACACGAGCAAGAGTTCCTTAAGGTAC 558
Db 181 AspGlnThrProArgValPheTyrLeuAlaGlyAsnProAspIleGluTyr 197
QY 559 CAGCAACAGCAGCAACAGCAGAGAGAGTTCATACCATAGCCCATACACCGCGCA 618
Db 198 -----ProGluThrMetGln 202
QY 619 AGTCAGCTAGCAGAGAGCGTGAATTTAGCCCTCGAGGACAGCAGCAGCGCAGAGAA 678
Db 203 GlnGlnGlnGlnLysSerHisGlyGlyArgLysGlnGlnHisGln 219
QY 679 CGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 738
Db 220 -----GlnGluGluGluGluGlySerValLeuSerGlyPheSerLysHis 236
QY 739 TTCCTGGAACAGCGCTTCAGGTTGACGACAGCAGATAGTGAACAACTTAAGAGCGAG 798
Db 237 PheLeuAlaGlnSerPheAsnThrAsnGlu---AspIleAlaGluLysLeu---Gln 253
QY 799 ACCGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 858
Db 254 SerProAspGluArgLysGlnIleValThrValGluGlyGlyLeuSerValIleSer 273
QY 859 CCAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 903
Db 274 ProLysTrpGlnGlnAspGluAspGluAspGluAspGluAspGluAspGluAspGlu 293
QY 904 -----GATCAATAT 912
Db 294 GlnIleProSerHisProProArgArgProSerHisGlyLysArgGluGlnAspGluAsp 313
QY 913 GAATACGATGAAGAGATAGA---AGGCGTGGCAGGGAAGCAGAGGC----- 957
Db 314 GluAspGluAspGluAspLysProArgProSerArgProSerGlnGlyLysArgGluGln 333
QY 957 ----- 957

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 8, 2005, 04:11:27 ; Search time 41.194 Seconds
(without alignments)

5523.384 Million cell updates/sec

Title: US-10-728-051-3

Perfect score: 2691

Sequence: 1 cggcagcaaccggagagaa.....ctccgagggtgtggttaa 1524

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 1027090

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10728051/runat_07042005_125046_14072/app.query.fasta_1.4757
-DB=Issued PatentsAA -QPM=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOPEXT=1 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10728051@cgn 1 122 @runat_07042005_125046_14072 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	811	30.1	492	4	US-09-462-720-2
2	293	10.9	141	4	US-09-645-593-12
3	227.5	8.5	85	4	US-09-645-593-10
4	215	8.0	165	4	US-09-645-593-11
5	184.5	6.9	489	4	US-09-424-283-3
6	173.5	6.4	96	4	US-09-645-593-9
7	172.5	6.4	626	4	US-09-106-872A-4
8	167	6.2	524	4	US-09-424-283-1
9	160	5.9	448	4	US-09-323-195A-18
10	158	5.9	523	4	US-09-323-195A-17
11	157.5	5.9	335	4	US-09-106-872A-17
12	156.5	5.8	587	1	US-09-955-905A-23

13	156	5.8	566	1	US-07-955-905A-2	Sequence 2, Appli
14	156	5.8	566	1	US-07-955-905A-22	Sequence 28, Appl
15	156	5.8	611	4	US-09-216-393B-81	Sequence 81, Appl
16	153.5	5.7	605	1	US-07-955-905A-24	Sequence 24, Appl
17	152	5.6	421	1	US-07-955-905A-27	Sequence 27, Appl
18	147.5	5.5	437	4	US-09-538-092-876	Sequence 876, App
19	147.5	5.5	614	5	PCT-US95-03236-21	Sequence 21, Appl
20	141.5	5.3	1898	1	US-08-056-200-94	Sequence 94, Appl
21	141.5	5.3	1898	2	US-08-800-644-94	Sequence 94, Appl
22	141.5	5.3	1898	4	US-09-538-092-1280	Sequence 1280, Ap
23	140.5	5.2	632	4	US-09-252-991A-25544	Sequence 25544, A
24	140	5.2	409	4	US-09-424-283-4	Sequence 4, Appli
25	139.5	5.2	481	4	US-09-252-991A-17707	Sequence 17707, A
26	139.5	5.2	1098	4	US-09-252-991A-20186	Sequence 20186, A
27	137	5.1	878	4	US-09-252-991A-31974	Sequence 31974, A
28	136	5.1	1250	1	US-08-441-139-9	Sequence 9, Appli
29	135.5	5.0	1178	4	US-09-543-681A-7819	Sequence 7819, Ap
30	135	5.0	444	4	US-09-424-283-2	Sequence 2, Appli
31	135	5.0	571	1	US-07-955-905A-25	Sequence 25, Appl
32	132	4.9	822	4	US-09-252-991A-21920	Sequence 21920, A
33	132	4.7	1093	4	US-09-252-991A-21827	Sequence 21827, A
34	130.5	4.8	648	4	US-09-252-991A-23682	Sequence 23682, A
35	130	4.8	553	4	US-09-370-767-43146	Sequence 43146, A
36	128.5	4.8	794	4	US-09-252-991A-28569	Sequence 28569, A
37	128	4.8	1971	4	US-09-914-272A-1	Sequence 1, Appli
38	128	4.8	1971	4	US-10-638-333-1	Sequence 1, Appli
39	127.5	4.7	405	4	US-09-949-016-11020	Sequence 11020, A
40	127.5	4.7	1043	4	US-09-252-991A-28885	Sequence 28885, A
41	127.5	4.7	1162	2	US-08-728-323A-2	Sequence 2, Appli
42	127.5	4.7	1162	3	US-09-298-568-2	Sequence 2, Appli
43	127.5	4.7	1162	4	US-09-410-399-2	Sequence 2, Appli
44	127.5	4.7	1162	4	US-09-894-273-2	Sequence 2, Appli
45	126.5	4.7	392	4	US-09-248-796A-16779	Sequence 16779, A

ALIGNMENTS

RESULT 1

US-09-462-720-2
; Sequence 2, Application US/09462720
; Patent No. 6617433
; GENERAL INFORMATION:
; APPLICANT: SOCIETE DES PRODUITS NESTLE
; APPLICANT: Maccacini, Pierre
; APPLICANT: Rogers, John
; TITLE OF INVENTION: COFFEE STORAGE PROTEINS
; FILE REFERENCE: 8265-303
; CURRENT APPLICATION NUMBER: US/09/462, 720
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: PCT/EP98/04038
; PRIOR FILING DATE: 1997-07-12
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Plant
; US-09-462-720-2

Alignment Scores:	2,81e-69	Length:	492
Pred. No.:	811.00	Matches:	182
Score:	52.42%	Conservative:	89
Best Local Similarity:	35.20%	Mismatches:	172
Query Match:	30.14%	Indels:	74
DB:	4	Gaps:	12

US-10-728-051-3 (1-1524) x US-09-462-720-2 (1-492)

Qy 1 CGGAGCAACCGGAG-----GAGAACGGCGCAGTTCAGCGCTCAATGCGCAG 51

Db 25 ArgProGlnProArgLeuArgGlyLysThrGlnCysAspIleGlnLysLeuAsnAlaGln 44

QY 52 AGACCTGACAAATCGCATTAATGAATCAGAGGCGGTTTACATTGAGACTTGGAAACCCCAACAC 111
Db 45 GluProSerPheArgPheProSerGluAlaGlyLeuThrGluPheTrpAspSerAsnAsn 64
QY 112 CAGGAGTTGAATCGCGCGCTCGCCCTCTCTCGCTTAGTCTCCGCGCGCAACGCCCTT 171
Db 65 ProGluPheGlyCysAlaGlyValGluPheGluArgAsnThrValGlnProLysGlyLeu 84
QY 172 CGTAGGCTTTCCTTCTCAATGCTCCCAAGAGATCTTCATCCAGCAAGAAAGGGATAC 231
Db 85 ArgLeuProHisTyrSerAsnValProLysPheValTyrValValGluGlyThrGlyVal 104
QY 232 TTTGGGTGATATCCCTGGTTCCTAGACACTATGAGAGCCTCACACAAAGTCTGT 291
Db 105 GlnGlyThrValIleProGlyCysAlaGluThrPheGlu-----SerGlnGlyGlu 121
QY 292 CGA-----TCTCAGTCCCAACACCAACAGCAGTCTCCAAGGAGAACACCA----- 339
Db 122 SerPheTrpGlyGlnGluGlnProGlyLysGlyGlnGluGlnGlnGlySer 141
QY 340 -----AGCCCAACAGCAACGA-----GATAGTCACCAAGGTGACCGT 378
Db 142 LysGlyGlnGluGlyArgGlnArgPheProAspArgHisGlnLysLeuArgArg 161
QY 379 TTCATGAGGTGATCTCATTCAGATTCCACCGGTGTTCTTCTGGCTCTACACAGAC 438
Db 162 PheGlnLysGlyAspValLeuLeuLeuLeuProGlyPheThrGlnTrpThrAsnAsp 181
QY 439 CACGACTGATGTGTGTCTTCTTCTTACTGACACCAACCAACAGCAACCAAGTTC 498
Db 182 GlyAspValProLeuValThrValAlaLeuLeuAspValAlaAsnGluAlaAsnGlnLeu 201
QY 499 GATCAGTTCCCGAGAGATCAATTTGCTGGGACACGAGCAAGATTCCTTAAGGTAC 558
Db 202 AspLeuGlnSerLysPhePheLeuAlaGlyAsnProGlnGlnGlyGlyLysGlu 221
QY 559 CAGCAACAAACGACAAAGCAGCAAGAGCTTACCATATAGCCCATACGCCCGCAA 618
Db 222 GlyHisGlnGlyGlnGlnGlnHisArg----- 231
QY 619 AGTCAGCTAGCAAGAGAGCGTGAATTTAGCCCTCGAGGACAGCAGCCGACAGAA 678
Db 231 ----- 231
QY 679 CGAGCAGCAGAGAGAGAAACGAAGTGGAAACATCTTCAGCGGCTTCACGCCGAG 738
Db 232 -----AsnIlePheSerGlyPheAspGln 240
QY 739 TTTCTGGAACAGCCTTCAGGTTGACGACAGACAGATAGTGCAAAACCTTAAGAGCGAG 798
Db 241 LeuLeuAlaAspAlaPheAsnVal---AspLeuLysIleIleGlnLysLeuLysGly--- 258
QY 799 ACCGAGATGAGAGAGAGAGCCATTGTGACAGTGAGGGA---GGCCTCAGAAATCTTG 855
Db 259 ---ProLysAspGlnArgGlySer-----ThrValArgAlaGluLysLeuGlnLeuPhe 275
QY 856 AGCCCAATGAGAAAGAGAGCTCCGACCAAGAGAGAAATACGATGAAGATGAATGAA 915
Db 276 LeuProGluTyrSerGluGlnValGlnGlnProGlnGln----- 288
QY 916 TAGCATGAGAGATAGAGGCGTGGCAGGGAACGAGGAGGAGGGAATGGTATTGAA 975
Db 289 ---GlnGlnGlnGlnGlnHisGlyValGlyArgGlyTrpArgSerAsnGlyLeuGlu 307
QY 976 GAGACCATCTGCACCCCAAGTGTAAAGAAACATTTGGTAGAAACAGATCCCTCGACATC 1035
Db 308 GluThrLeuCysThrValLysLeuSerGluAsnIleGlyLeuProGlnGluAlaAspVal 327
QY 1036 TACAACCTCAAGCTGCTTCAAACTGCCACGATCTCAACTTCTAATCTAGG 1095
Db 328 PheAsnProArgAlaGlyArgIleThrValAsnSerGlnLysIleProIleLeuSer 347
QY 1096 TGGCTTGACCTAGTGTGAATGTGAATCTCTACAGGAATGCTATGTTTCTGCTCAC 1155

Db 348 SerLeuGlnLeuSerAlaGluArgGlyPheLeuTyrSerAsnAlaIlePheAlaProHis 367
QY 1156 TACAACACCAACGACACACAGCATCATATATCGATTGAGGGGACGGGCTCACGTGCAAGTC 1215
Db 368 TrpAsnIleAsnAlaHisAsnAlaLeuTyrValIleArgGlyAsnAlaArgIleGlnVal 387
QY 1216 GTGCACAGCAACGGCAACAGAGTGATACGACGAGGAGCTTCAAGAGGGTCACTGCTTGTG 1275
Db 388 ValAspHisLysGlyAsnLysValPheAspAspValLysGlnGlyGlnLeuIleIle 407
QY 1276 GTGCCACAGAACTTCGCTCGCTGGAAGTCCACAGAGCGAGAACTTCGAATACGTGGCA 1335
Db 408 ValProGlnTyrPheAlaValIleLysLysAlaGlyAsnGlnGlyPheGluTyrValAla 427
QY 1336 TTCAGACACAGCTCAAGGCCAGCATAGCCCACTCGCGGTGCAAACTCCGTCATAGAT 1395
Db 428 PheLysThrAsnAspAsnAlaMetIleAsnProLeuValGlyArgLeuSerAlaPheArg 447
QY 1396 AACCTGCGGAGGAGGTGTGCAATTCATATGCGCTCCAAAGGAGCAGGCAAGGCAG 1455
Db 448 AlaIleProGluGluValLeuArgSerSerPheGlnIleSerGluGluAlaGluGlu 467
QY 1456 CTTAAGAACACACACCCCTTCAAGTTCTTCTGTTCCACCGTCTCAGCAGTCT 1506
Db 468 LeuLysTyrGlyArgGlnGluArgLeuLeu-----SerGluGlnSer 482

RESULT 2
US-09-645-593-12
; Sequence 12, Application US/09645593
; Patent No. 6777591
; GENERAL INFORMATION:
; APPLICANT: Chaudhary, Sarita
; APPLICANT: van Rooijen, Gifs
; APPLICANT: Moloney, Maurice
; APPLICANT: Singh, Surinder
; TITLE OF INVENTION: Flax Seed Specific Promoters
; FILE REFERENCE: 9369-151
; CURRENT APPLICATION NUMBER: US/09/645,593
; CURRENT FILING DATE: 2000-08-25
; PRIOR APPLICATION NUMBER: US 60/151,044
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: US 60/161,722
; PRIOR FILING DATE: 1999-10-27
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 12
; LENGTH: 141
; TYPE: PRT
; ORGANISM: Linum usitatissimum
US-09-645-593-12

Alignment Scores:
Pred. No.: 1,21e-19 Length: 141
Score: 293.00 Matches: 55
Percent Similarity: 72.73% Conservative: 25
Best Local Similarity: 50.00% Mismatches: 30
Query Match: 10.89% Indels: 0
Gaps: 0
DB:

US-10-728-051-3 (1-1524) x US-09-645-593-12 (1-141)

QY 1138 GCATTGTTTGTCTGCTCACTACACCAACGACAGCATCATATATCGATTGAGGGA 1197
Db 2 AlaIleArgLeuProHisTrpAsnIleAsnAlaHisSerIleValTyrAlaIleArgGly 21
QY 1198 CGGCTCAGCTGCAAGTCTGGGACGACCAACGCGCAACAGAGTGTTACACGAGAGCTTCAA 1257
Db 22 GlnAlaArgValGlnIleValAsnGluGluGlyAsnSerValPheAspGlyValLeuGln 41
QY 1258 GAGGCTCAGCTGCTGTTGGTGGCCACAGAACTTCGCCGTCCGTGGAAAGTCCCGACGAG 1317
Db 42 GluGlyGlnValValThrValProGlnAsnPheAlaValValLysArgSerGlnSerGlu 61

```
QY 1318 AACTTGAATACGTGGCAATTCAGACAGACTCAAGGCCAGCATAGCCAACTCGCGGT 1377
Db 62 ArgPheGluTrpValAlaPheLysThrAsnAspAlaMetValAsnSerLeuAlaGly 81
QY 1378 GAAAACTCCGCTAGATAACCTGCCGAGAGGTGGTTCGAAATTCATATGCGCTCCAA 1437
Db 82 ArgThrSerAlaValArgAlaIleProAlaAspValLeuAlaAsnAlaIleTrpArgValSer 101
QY 1438 AGGAGCAGGCAAGCGCAGCTTAAGACAAC 1467
Db 102 ProGluGluAlaArgArgValLysPheAsn 111

RESULT 3
US-09-645-593-10
; Sequence 10, Application US/09645593
; Patent No. 6777591
; GENERAL INFORMATION:
; APPLICANT: Chaudhary, Sarita
; APPLICANT: van Rooijen, Gijb
; APPLICANT: Moloney, Maurice
; APPLICANT: Singh, Surinder
; TITLE OF INVENTION: Flax Seed Specific Promoters
; FILE REFERENCE: 9369-151
; CURRENT APPLICATION NUMBER: US/09/645,593
; CURRENT FILING DATE: 2000-08-25
; PRIOR FILING DATE: 1999-08-27
; PRIOR FILING DATE: 1999-10-27
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 10
; LENGTH: 85
; TYPE: PRT
; ORGANISM: Linum usitatissimum
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (59)
; OTHER INFORMATION: n is any amino acid
; NAME/KEY: SITE
; LOCATION: (62)
; OTHER INFORMATION: n is any amino acid
; NAME/KEY: SITE
; LOCATION: (78)
; OTHER INFORMATION: n is any amino acid
; NAME/KEY: SITE
; LOCATION: (82)
; OTHER INFORMATION: n is any amino acid
; OTHER INFORMATION: n is any amino acid
US-09-645-593-10

Alignment Scores:
Pred. No.: 1,9e-13 Length: 85
Score: 227.50 Matches: 46
Percent Similarity: 61.22% Conservative: 14
Best Local Similarity: 46.94% Mismatches: 25
Query Match: 8.45% Indels: 13
DB: 4 Gaps: 2

US-10-728-051-3 (1-1524) x US-09-645-593-10 (1-85)
QY 220 GGAAGGGATACCTTGGTGTGATATCCCTGGTGTCTTACACACTATGAAGACCTCAC 279
Db 1 GlyArgGlyValThrGlyIleMetPhePro**CysProGluThrPheGluGluSerGln 20
QY 280 ACACAAGGTGCTGATCTCAGTCCCAAGACCACCAAGACGCTCTCCAAGGAGAGACCAA 339
Db 21 GlnGlnGlyGlnGlnGlyGln-----GlnGlySerSerGln 32
QY 340 AGCCACAGCAACAGAGATAGTCACAGAGAGGTGCACCGTTTCGATGAGGGTGATCTCAT 399
Db 33 -----AspGlnHisGlnIleArgPheArgPheArgGluGlyAspValIle 47
```

```
QY 400 GCAGTTCCACCGGTGCTTCTGGCTCTACAACGACACGACACTGATGTGTGCT 459
Db 48 AlaValProAlaGlyValAlaHisTrpSerTyrAsnAspGlyAsnGluProValMetAla 67
QY 460 GTTCTCTTACTCACACCAACAACAACGACGACGCTTTCATCAGTTCCCCAGG 513
Db 68 IleValValHisAspThrSerSerHisLeuAsnGlnLeuAspAsnProArg 85

RESULT 4
US-09-645-593-11
; Sequence 11, Application US/09645593
; Patent No. 6777591
; GENERAL INFORMATION:
; APPLICANT: Chaudhary, Sarita
; APPLICANT: van Rooijen, Gijb
; APPLICANT: Moloney, Maurice
; APPLICANT: Singh, Surinder
; TITLE OF INVENTION: Flax Seed Specific Promoters
; FILE REFERENCE: 9369-151
; CURRENT APPLICATION NUMBER: US/09/645,593
; CURRENT FILING DATE: 2000-08-25
; PRIOR FILING DATE: 1999-08-27
; PRIOR FILING DATE: 1999-10-27
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 11
; LENGTH: 165
; TYPE: PRT
; ORGANISM: Linum usitatissimum
; US-09-645-593-11

Alignment Scores:
Pred. No.: 4,29e-12 Length: 165
Score: 215.00 Matches: 62
Percent Similarity: 43.06% Conservative: 28
Best Local Similarity: 29.67% Mismatches: 69
Query Match: 7.99% Indels: 50
DB: 4 Gaps: 6

US-10-728-051-3 (1-1524) x US-09-645-593-11 (1-165)
QY 517 TTCAATTTGGCTGGAAACGAGAGAGCTTCTTAAGTACCAGCAACAAGACAGACAA 576
Db 2 PheTyrLeuAlaGlyAsnProArgAspGluPheGluGlnSerGlnGlnGlyArgLeu 21
QY 577 AGCAGACGAGAGAGCTTACCATATAGCCCATACAGCCGCAAGTACGCTAGACAAGAA 636
Db 22 SerArgGlyGluSer----- 26
QY 637 GAGCGTGAATTTAGCCCTCGAGGACAGACAGCCGCGAGAGAACCGAGCAGCAAGAAGAA 696
Db 27 -----GluGlyGlyArgGlyArgGluProLeuGlnProAlaThr 40
QY 697 GAAAACGAGGTGGAACATCTTCAGCGGCTTCACGCCGAGTTCCTGGAAACAGCCTTC 756
Db 41 ThrSerSer-----CysGlyIleAspSerLysLeuIleAlaGluAlaPhe 55
QY 757 CAGGTTACGACAGACAGATAGTGCAGAACCTTAAGAGCCGAGACCGAGAGTGAAGAGAG 816
Db 56 AsnValAspGlu---AsnValAlaArgArgLeu-----GlnSerGluAsnAspAsnArg 72
QY 817 GGAGCCATTTGACAGTGGAGGGGCGCTCAGAAATCTTGAGCCAGCATAGAAAGAGAGACGT 876
Db 73 GlyGlnIleValArgValGluGlyGluLeuAspIleValArgProProThrSerIleGln 92
QY 877 GCCGACGAGAGAGAGAAATACGATGAATATGATATACGATGAAGAGGATAGAAGG 936
Db 93 GluGluSerGlnGluGln----- 98
QY 937 CTGGCGAGGGGAGACGAGCAGCGGGG-----AATGTTATTGAGACGATC 984
```

Db 99 -----GlycylArgGlyGlyArgTyrSerAsnGlyValGluGluThrPhe 115
QY 985 TGCACCGCAAGTCCTAAAAGACATTGGTAGAAACAGATCCCTGACATCTACAACCT 1044
Db 116 CysSerMetArgLeuLeuGluAsnIleGlyAspProSerArgAlaAspIlePheThrPro 135
QY 1045 CAAGCTGGTTCACTCAAACTGCCACGATCTCAACCTTCTTAATCTAGGTGGCTTGA 1104
Db 136 GluAlaGlyArgValArgSerLeuAsnSerHisAsnLeuProValIleGlnTrpIleGln 155
QY 1105 CTTAGTGTGAATATGAAATCTCTAC 1131
Db 156 LeuSerAlaGluArgGlyValLeuTyr 164

RESULT 5
US-09-424-283-3
; Sequence 3, Application US/09424283
; Patent No. 6437219
; GENERAL INFORMATION:
; APPLICANT: Grimes, et al.
; TITLE OF INVENTION: Sucrose binding proteins
; FILE REFERENCE: 4630-50206
; CURRENT APPLICATION NUMBER: US/09/424,283
; PRIOR FILING DATE: 1999-11-19
; PRIOR FILING DATE: PCT/US98/10465
; PRIOR FILING DATE: 1998-05-21
; PRIOR FILING DATE: US 60/047,568
; PRIOR FILING DATE: 1997-05-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 489
; TYPE: PRT
; ORGANISM: Glycine max
US-09-424-283-3

Alignment Scores:
Pred. No.: 6.51e-09 Length: 489
Score: 184.50 Matches: 99
Percent Similarity: 34.19% Conservative: 87
Best Local Similarity: 18.20% Mismatches: 187
Query Match: 6.86% Indels: 171
DB: 4 Gaps: 17

US-10-728-051-3 (1-1524) x US-09-424-283-3 (1-489)
QY 1 CGCAGCAACCGGAGGAGACGGTGCAGTTCAGCGCTCAATGCGCAGACCT--- 57
Db 80 ArgGluLysGluGluGluHisGlnGluGlnHisGluGluGluGluGluAsnProTyr 99
QY 58 -----GACAATCGCATTGAATCAGAGCGCGGTACATTGAG 93
Db 100 ValPheGluGluAspLysAspPheSerThrArgValGluThrGluGlySerIleArg 119
QY 94 ACTTGAACCCCAACACACGAGGTTCGAA---TGCGCGCGGTGCGCCCTCTCTCGCTTA 150
Db 120 ValLeuLysLysPheThrGluLysSerLysLeuLeuGlnGlyIleGluAsnPheArgLeu 139
QY 151 GTCCTC-----CGCGCAACGCCCTTCGTAGGCTTCTACTCCATGCTCCCCAG 201
Db 140 AlaIleLeuGluAlaArgAlaHisThrPheValSerProArgHisPheAspSerGluVal 159
QY 202 GAGATCTTCATCCAGCAAGGAGGGATCTTTGGGTGATATTCCTGCTGTGCTCTAGA 261
Db 160 ValLeuPheAsnIleLysGlyArgAlaValLeuGlyLeuVal----- 173
QY 262 CACTATGAGAGCCTCACACAAAGGTCGTGATCTCAGTCCCAACACCAACCAAGAGCT 321
Db 173 ----- 173
QY 322 CTCACGAGNAGACCAAGCCACACGAGATAGTCACCAAGAGGTGCACCGTTTC 381
Db 174 -----ArgGluSerGluThrGluLysIleThrLeu 183

QY 382 GATGAGGTGATCTCATTGCGATTTCCACCGGTGTGCTTTCTGGCTCTACAACGACAC 441
Db 184 GluProGlyAspMetIleHisIleProAlaGlyThrProLeuTyrIleValAsnArgAsp 203
QY 442 GACACTGATGTGTGCTGCTTTCTTACTGACACCAACAACAGACCAACCGCTTGTAT 501
Db 204 GluAsnGluLysLeuLeuAlaMetLeu----- 213
QY 502 CAGTTCCTCCAGGAGATTCAATTGCTGGGAACACGGAGCAAGAGTCTTAAAGGTACCAG 561
Db 213 ----- 213
QY 562 CAACAAAGACAGCAAGACGACGAGAAGCTTTACCATATAGCCCATACAGCCCGCAAGT 621
Db 214 -----HisIleProValSer 218
QY 622 CAGCCTAGACAAGAGAGCGTGAATTTAGCCCTCGAGGACACGACGCGGACAGAACGA 681
Db 219 ThrProGlyLysPheGluGluPheGlyProGlyGlyArg----- 232
QY 682 GCAGGACAAGAGAAGAAAACGAAGTGGAAACATCTTCAGCGGCTTCACGCGGATTC 741
Db 233 -----AspProGluSerValLeuSerAlaPheSerTrpAsnVal 245
QY 742 CTGGAACAAGCCTTCAGGTTGACGACACAGATAGTGCAAAACCTAAGAGCGGACGCC 801
Db 246 LeuGlnAlaAlaLeuGlnThrPro-----LysGlyLysLeu 257
QY 802 GAG-----AGTGAAGAAGAGGAGCCATTGTGACAGTG---AGGGGAGCCCTC 846
Db 258 GluArgLeuPheAsnGlnGlnAsnGluGlySerIlePheLysIleSerArgGluArgVal 277
QY 847 AGAATCTTGAGCCACAGATAGAAAGAGACGTGCGCAGCAAGAGAGAAATACATGAAGAT 906
Db 278 ArgAlaLeuAlaProThrLysLys-----SerSer 287
QY 907 GAATATGAATACGATGAAGAGATAGAAGCGCTGCGAGGGGAAGCAGAGCGAGGGGAT 966
Db 288 TrpTrpProPheGlyGlyGluSerLys----- 296
QY 967 GGTATTGAAGACACGATCTGCACCCGACAGTGTCTAAAGAACATTGGTAGAACACAGATCC 1026
Db 297 -----AlaGlnPheAsnIlePheSerLysArg--- 305
QY 1027 CCGTACATCTACAACCTCAAGCTGTTCACTCAAAAACCTGCCAACGATCTCAACCTCTCA 1086
Db 306 ProThrPheSerAsnGlyTyrGlyArgLeuThrGluValGlyProAspAspGluLysSer 325
QY 1087 ATACTTAGTGGCTGGACCTAGTGTGAATATGGAATCTCTACAGGAATGCATTGTTT 1146
Db 326 TrpLeuGlnArgLeuAsnLeuMetLeuThrPheThrAsnIleThrGlnArgSerMetSer 345
QY 1147 GTCGCTCACTACAACCAACGACGACATCATATATCGATTGAGGAGCGGCTCAC 1206
Db 346 ThrIleHisThrAsnSerHisAlaThrLysIleAlaLeuValMetAspGlyArgGlyHis 365
QY 1207 GTGCAAGTC-----GTGCACAGCAACGCAACAGAGGTGTACACGAG----- 1248
Db 366 LeuGlnIleSerCysProHisSerHisMetSerSerArgSerAspSerLysHisAspLysSer 385
QY 1249 -----GAGCTTCAAGAGGTCACGTGCTTGTGTCGCCACAG 1284
Db 386 ProSerTyrHisArgIleSerAlaAspLeuLysProGlyMetValPheValProPro 405
QY 1285 AACTTCGCC---GTCGCTGMAAGTCCAGAGCGAAGCTTCGAATACGTGGCATTCAG 1341
Db 406 GlyHisProPheValThrIleAlaSerAsnLysGluAsnLeuLeuIleCysPheGlu 425
QY 1342 ACAGACTCAAGGCCCGCCAGCATAGCC---AACCTCGCGGTGAAACCTCCGTCATA----- 1392
Db 426 ValAsnValArgAspAsnLysLysPheThrPheAlaGlyLysAspAsnIleValSerSer 445

01018	AA	CAGATGCCCTGACATCTACAAACCTCAAGCTGGGTCACTCAAAAACTGCCACAGATCTC	107
331	----	-----ValGlyProAspAspAspGluLysSerTrpLeuGlnArgLeu	344
1078	AA	CTTCTTAATACTTAGTGGCTTGGACCTAGTGTGAATATGTAATCTCTACAGGAAT	1137
345	Asn	LeuMetLeu-----ThrPheThrAsnIleThrGlnArg	356
1138	GC	ATTGTTTGTCTCACTCAACACACGACGACATCATATATCGATTCAGGCGGA	1197
357	Ser	MetSerThrIleHisItyrAsnSerHisAlaThrLysIleAlaLeuValIleAspGly	376
1198	CG	GCTCAGCTGCAAGTC-----GTGGACAGCAACGGCAACGAGAGTGTC	1242
377	Arg	GlyHisLeuGlnIleSerCysProHisMetSerSerArgSerSerHisSerLysHis	396
1243	GAC	GAG-----GAGCTTCAAGAGGCTCAGCTGCTT	1272
397	Asp	LysSerSerProSerTyrHisArgIleSerSerAspLeuLysProGlyMetValPhe	416
1273	GT	GCTGCCACAGAACTTCGCC-----GTCGTGGAAAGTCCCAGACGACGAACCTTCGAATAC	1329
417	Val	ValProProGlyHisProPheValThrIleAlaSerAsnLysGluAsnLeuMet	436
1330	GT	GCAATTCAAGACAGACTCAAGCCGACGATGCC-----AACCTCGCCGGTGAAACTCC	1386
437	Ile	CysPheGluValAlaAlaArgAspAsnLysLysPheThrPheAlaGlyLysAspAsn	456
1387	GT	CATAGAT-----AACCTGCCGGAGGAGGTG	1413
457	Ile	ValSerSerLeuAspAsnValAlaLysGluLeuAlaPheAsnTyrProSerGluMet	476
1414	GT	TGCAAAATTCATATGGCTCCAAAGGAGCAGGCAAGCAGCTT-----	1458
477	Val	AsnGlyValPheLeuLeuGlnArgPheLeuGluArgLysLeuIleGlyArgLeuTyr	496
1459	-----	-----AAGAACAAACAACCCCTTCAAGTTCTTTCGTTCCACCCGTCT	1497
497	His	LeuProHisLysAspArgLysGluSerPhePheProPheGluLeuProArgGlu	516
1498	CAC	AGTCTCCGAGGCTGTGGCT	1521
517	Glu	ArgGlyArgAlaAspAla	524
RESULT 9			
US-09-323-195A-18			
; Sequence 18, Application US/09323195A			
; Patent No. 6462257			
; GENERAL INFORMATION:			
; APPLICANT: Pullman, Gerald			
; APPLICANT: Cairney, John			
; APPLICANT: Perrera, Nanjan			
; TITLE OF INVENTION: VANICILIN-LIKE SEED STORAGE PROTEIN GENE PROMOTER AND			
; TITLE OF INVENTION: METHODS OF USING THE SAME			
; FILE REFERENCE: IPST0009			
; CURRENT APPLICATION NUMBER: US/09/323,195A			
; CURRENT FILING DATE: 1999-06-01			
; NUMBER OF SEQ ID NOS: 19			
; SOFTWARE: Patentin Ver. 2.0			
; SEQ ID NO 18			
; LENGTH: 448			
; TYPE: PRT			
; ORGANISM: Picea glauca			
US-09-323-195A-18			
Alignment Scores:			
Pred. No.:	1.43e-06	Length:	448
Score:	160.00	Matches:	103
Percent Similarity:	33.59%	Conservative:	70
Best Local Similarity:	20.00%	Mismatches:	172
Query Match:	5.95%	Indels:	170
DB:	4	Gaps:	20

DB:

Alignment Scores:	
Pred. No.:	1.43e-06
Score:	160.00
Percent Similarity:	33.59%
Best Local Similarity:	20.00%
Query Match:	5.95%
DB:	4
Length:	448
Matches:	103
Conservative:	70
Mismatches:	172
Indels:	170
Gaps:	20

```
US-10-728-051-3 (1-1524) x US-09-323-195A-18 (1-448)
QY 4 CAGCAACCGGAGGAGGAGCGTCCAGTTCAGCGCTCAATGCGCAGAGACCTGACAAT 63
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
46 GluGluArgGluGluAsnProTyrValPhe-----HisSerAspSerPheArgThr 62
QY 64 CGCATTGAATCAGAGGCGGTTTACATTGAGACTTGAACCCCAACACAGGAGTTTCGAA 123
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
63 ArgAlaSerSerGluAlaGlyGluIleArgAlaLeu---ProAsnPheGlyGluValSer 81
QY 124 -----TGGCGCGCGTCCGCTCTCTCGTTAGTC-----CTCGCGCGCAACGCC 168
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
82 GluLeuLeuGluGlyIleArgLysPheArgValThrCysIleGluMetLysProAsnThr 101
QY 169 CTTTCGTAGGCTTTCTACTCCAATGCTCCCGAGGATCTTCATCCAGCAAGGAGGGA 228
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
102 ValMetLeuProHisTyrIleAspAlaThrTrpIleLeuTyrValThrArgGlyArgGly 121
QY 229 TACTTTGGGTGTATATTCCCTGGTTGCTCTAGACACTATGAAGAGCCTCACACAAAGT 288
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
122 TyrIleAlaTyrVal----- 126
QY 289 CGTCGATCTCAGTCCCAAGACCAACAGAGCTCTCAAGGAGAGAACCAAGCCAAAG 348
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
127 -----HisGlnAsnGluLeuValLysArgLysLeu----- 136
QY 349 CAACGAGATAGTACCAGAAAGTGCCCGTTTCGATGAGGGTGATCTCATTCAGTCC 408
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
137 -----GluGluGlyAspValPheGlyValPro 145
QY 409 ACCGCTGTTGTTCTTGGCTCTACACGACACAGCACTGATGTTGTTGTTCTCTT 468
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
146 SerGlyHisThrPheTyrLeuValAsnAspAspHisAsnThrLeuArgIleAla--- 164
QY 469 ACTGACACCAACAAACAGCAACAGCTTGATCAGTTCCTCCAGGAGATTCATTTGGCT 528
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
164 ----- 164
QY 529 GGGAACACGGAGCAGAGTTCTTAAGTACCAGCAACAAAGCAGACAAAGCAGCAAG 588
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
164 ----- 164
QY 589 AGCTTACCATTATAGCCCATACAGCCCGCAAGTACAGCTAGACAAAGAGCGGTGATTT 648
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
165 -----SerLeuValArgProValSerThrValArgGlyGlyGlnProPhe 180
QY 649 AGCCCTCGAGGACAGCACGCCCGCAGAACAGCAGGAGGACAGAAAGAAACGAAGT 708
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
181 TyrValAlaGly-----GlyArgAsnProGln----- 189
QY 709 GGAACATCTTCAGCGCTTCAGCGGAGTTCTCGAACCAAGCTTCAGGTTGACGAC 768
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
190 -----ThrValTyrSerAlaPheSerAspValLeuGluAlaAlaPheAsnThrAsnVal 208
QY 769 AGACAGATAGTCAAAACCTTAAGAGCGAGACCGCAGAGTGAAGAGAGGAGCCATTGTG 828
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
209 GlnGlnLeuGluArgIlePheGlyGly-----HisLysSerGlyValIleIle 224
QY 829 ACAGTGGGGAGGCTCAGATCTTGAGCCCGCAGATAGAAAGAGAGCTGCCGAGAGAA 888
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
225 -----HisAlaAsnGluGlu 229
QY 889 GAGGAATACGATCAAGATGAATATGATGATGAAGAGGATAGAGGCGTGGCAGGGA 948
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
230 -----GlnIleArgGluMetMetArgLysArgGly-----Phe 240
QY 949 AGCAGAGCGGAGGGAATGATTAAGAGAGACGATCTGCACCGCAAGTCTCAAAAGAAC 1008
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
241 SerAlaGlySerMetSerAlaProGluHis-----ProLysProPhe 254
QY 1009 ATTGGTAGAACAGATCCCTGATCTACACCTCAAGCTGGTTCACTCAAACTGCC 1068
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
255 AsnLeuArgAsnGlnLysProAspPheGluAsn---GluAsnGlyArgPheThrIleAla 273
QY 1069 AACGATCTCAACTTCTAATAGTCTAGGCTGGCTTGGACCTAGTCTGAATATGGAATCTC 1128
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
274 GlyProLysAsnTyrProPheLeuAspAlaLeuAspValSerValGlyLeuAlaAspLeu 293
QY 1129 TACAGGAATGCATGTTTGTGCTCCTACTACAACCAACGACACAGCAGCATATATCGA 1188
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
294 AsnProGlySerMetThrAlaProSerLeuAsnSerLysSerThrSerIleGlyIleVal 313
QY 1189 TTGAGGGGACGGCTCAGTCAAGTCGTG----- 1218
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
314 ThrAsnGlyGluGlyArgIleGluMetAlaCysProHisLeuGlyGlnHisGlyTrpSer 333
QY 1219 -----GACGCAACCGCAACAGAGTGTACGACGAGAGCTT 1254
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
334 SerProArgGluArgLysAspGlnAspIleThrTyrGlnArgValTrp---AlaLysLeu 352
QY 1255 CAAGAGGTCACGTGCTTGTGTCACAGAACTTCGCGCTGCTGGA---AAGTCCCAAG 1311
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
353 ArgThrGlySerValTyrIleValProAlaGlyHisProIleThrGluIleAlaSerThr 372
QY 1312 AGCGAGAATCTCGAATACGTGGCATTCAAGCAGACTCAAGCCCGCAGCATAGCCAAC--- 1368
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
373 AsnSerArgLeuGlnIleLeuTrpPheAspLeuAsnThrArgGlyAsnGluArgGlnPhe 392
QY 1369 CTCGCGGTGAAACTCCGCTCATAGTAACTCCGCGAGAGGTGTTGCAAAATTCATAT 1428
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
393 LeuAlaGlyLysAsnValLeuAsnThrLeuGluArgGluIle----- 407
QY 1429 GGCCTCCAAAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1473
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
408 -----ArgGlnLeuSerPheAsnValPro 415

RESULT 10
US-09-323-195A-17
; Sequence 17, Application US/09323195A
; Patent No. 6462257
; GENERAL INFORMATION:
; APPLICANT: Pullman, Gerald
; APPLICANT: Cairney, John
; APPLICANT: Ferrera, Ranjan
; TITLE OF INVENTION: VICILIN-LIKE SEED STORAGE PROTEIN GENE PROMOTER AND
; TITLE OF INVENTION: METHODS OF USING THE SAME
; FILE REFERENCE: IPST0009
; CURRENT APPLICATION NUMBER: US/09/323,195A
; CURRENT FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patent in ver. 2.0
; SEQ ID NO 17
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Pinus taeda
US-09-323-195A-17

Alignment Scores:
Pred. No.: 2,41e-06 Length: 523
Score: 158.00 Matches: 107
Percent Similarity: 32.67% Conservative: 74
Best Local Similarity: 19.31% Mismatches: 190
Query Match: 5.87% Indels: 183
DB: 4 Gaps: 20

US-10-728-051-3 (1-1524) x US-09-323-195A-17 (1-523)
QY 10 CCGGAGGAGAACCGTGGCAGTTCAGCGCTCAATGCCAGAGACCTGACAAT----- 63
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
37 ProGluAspHisGlyArgGlyHisGlnArgGluGluGluGluGluGluGluGluGluGlu 56
QY 64 -----CGCATTGAATCAGAGCGGCGGTTTACATTGAGACT 96
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
57 ValPheHisSerAspArgPheArgMetArgAlaSerSerAspAlaGlyGluIleArgAla 76
```


97 TGGAAACCCCAACACAGGAGTTGCGA-----TGGCCCGCGCTGCCCTCTCTCGCTTA 150
Db Leu---ProAsnPheGlyGluAlaSerGluLeuLeuGluGlyIleSerIleSfyrA-gVal 95
151 GTC-----CTCCGCCCAACCGCCTTCTGAGGCTTCTACTCCAATGTCTCCCGAG 201
Db ThrCysIleGluMetArgProAsnThrValMetLeuProHisIleSfyrLeuAspAlaThrTrp 115
202 GAGATCTTCATCCAGCAAGAGGGGATCTTTGGGTGGATATTCCTGGTGTCTCTAGA 261
Db IleLeuTyrValThrGlyGlyA-gGlyTyrIleAlaTyrVal----- 129
262 CACTATGAAGAGCTCACACACAGAGTCTCGATCTCAGTCCCAAGACACCAACAGAGCT 321
Db 130-----HisGlnAsnGluLeuValIleSfyrA-glys 138
322 CTCAAGGAGAGAACCAAGCAACAGCAACAGAGATAGTACCAGAGGTGCACCGCTTC 381
Db 139 Leu----- 139
382 GATGAGGTGATCTCATGTCAGTTCCACCGGTGTTGCTTTCTGGCTCTACACGACCAC 441
Db 140 GluGluGlyAspValPheGlyValProSerGlyHisThrPheTyrLeuValAsnAsp 159
442 GACACTGATCTGTGCTGTTCTCTTACTGACACCAACCAACAGACACACAGCTTGAT 501
Db 160 AspHisAsnSerLeuArgIleThr----- 167
502 CAGTTCCTCCAGGAGATTCAATTTGGCTGGGAACACGAGCAAGATTCTTAAGTACCAG 561
Db 168-----SerLeuLeuA-gThrVal 173
562 CACAAAGCAGACAAGCAGACGAGAGAGTTCATATAGCCCATACAGCCCGCAAGT 621
Db 174 SerThrMetArgGlyGlu-----TyrGluProTyr----- 183
622 CAGCCTAGACAAGAGCGGTGAATTTAGCCCTCGAGGACAGCAGCCGCGAGAACGA 681
Db 184-----TyrValAla 186
682 GCAGGACAAGAGAAAGAGAGTGGAAACATCTTCAGCGCTTCACGCGGAGTTC 741
Db 187 GlyIleArgAsnProGlu-----ThrValTyrSerAlaPheSerAspAspVal 202
742 CTGGAACAACCTTCCAGTTGACGACGACAGATAGTCAAAACCTTAAGAGCGAGACC 801
Db 203 LeuGluAlaAlaPheAsnThrAsn----- 210
802 GAGAGTGAAGAGGAGGCATTTGTACAGTGGGGGCGCTCAGAACTTTGAGCCCA 861
Db 211-----ValIleGluAlaArgThrHisPheProVal----- 220
862 GATAGAAAGAGAGCTGCCGAGAGAGAGATACGATGAAGATGAATATGAATACGAT 921
Db 221-----HisIleGluArgGluSerTyrSerMetAla 230
922 GAAGAGATAGAGGCTGGCAGGGGAGAGCAGAGGCGAGGGGAATGGTATTTAGAGAGC 981
Db 231 AsnGluGluGlnIleArg---GluMetLeuArgIleSfyrPheSerAlaGluSerMet 249
982 ATCTGACCCCAAGTGTAAAGAACATTTGGTAGAACAGATCCCTGTGATCTACAC 1041
Db 250 SerAlaSerGluHisProIleSfyrProPheAsnLeuArgAsnGlnIleSfyrPheGluAsn 269
1042 CCTCAAGCTGTTTCACTCAAACTGCCAAGCTCTCAACCTTCTAATACCTTAGTGCTT 1101
Db 270 AsnAsnGlyArgPheThrArgAlaGlyProAsnGluAsnProLeu---LeuAspAlaVal 288
1102 GGACCTAGTCTGAATATGAAATCTCTACAGGAATGCATTTGTTGCTGCTCACTACAC 1161
Db 289 AspValThrAlaGlyPheGlyValLeuAsnProGlyThrMetThrAlaProSerHisAsn 308
1162 ACCAACCCACACATCATATATCGATTGAGGGGACGGCTCACGTGCAAGTCTG--- 1218

Db 309 ThrLysAlaThrSerIleAlaIleValThrGlnGlyGluGlyArgIleGluMetAlaCys 328
Qy 1219-----GACAGCAAC 1227
329 ProHisLeuGlyGlnHisGlyTyrSerSerArgGluLysGlyAspGlnGluIleAsn 348
Qy 1228 GGCACAGAGTGTACACGAGGAGCTTCAAGAGGTCACGCTTGTGTGGTCCACAGAAC 1287
Db 349 TyrGlnArgVal---ArgAlaArgLeuA-gThrGlyThrValTyrValProAlaGly 367
Qy 1288 TTCCGCGTCTGCTGAAAGTCCAGAGCGAG---AACTTCGAATACGTGGCATTCAAGACA 1344
Db 368 HisProIleThrGluIleAlaCysThrGluGlyHisLeuGluIleLeuTrpPheAspIle 387
Qy 1345 GACTCAAGGCCCAAGATAGCAAC---CTCGCGGTGAAACCTCCGTATAGATAACCTG 1401
Db 388 AsnThrSerGlyAsnGluArgGlnPheLeuAlaGlyLysTyrAsnValLeuGlnThrLeu 407
Qy 1402 CCGGAGAGGTGTTGCAAAATTCATATGCTCCAAAGG-----GACCAG 1446
Db 408 GluLysGluValArgGlnIleSerPheAsnIleProArgGlyGluGluLeuAspGluVal 427
Qy 1447 GCAAGCGAGCTTAAGAAC-----AACAAAC 1470
Db 428 LeuArgArgGlnLysAspGlnValIleLeuArgGlyProGlnMetGlnArgAspGlu 447
Qy 1471 CCCTTCAAGTCTTCTTCCACCGCTCTCAGCAGCTCTCCGA 1510
Db 448 -ProArgSerSerIleHisAlaIleIleAlaAlaArg 460
RESULT 11
US-09-106-872A-17
; Sequence 17, Application US/09106872A
; Patent No. 6486311
; GENERAL INFORMATION:
; APPLICANT: Burks Jr., A. Wesley
; APPLICANT: Stanley, J. Steven
; APPLICANT: Cockrell, Gael
; APPLICANT: King, Nina E.
; APPLICANT: Sampson, Hugh A.
; APPLICANT: Helm, Ricki M.
; APPLICANT: Bannon, Gary A.
; TITLE OF INVENTION: Peanut Allergens and Methods
; FILE REFERENCE: HS 103 CIP
; CURRENT APPLICATION NUMBER: US/09/106, 872A
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: PCT/US96/15222
; PRIOR FILING DATE: 1996-09-23
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Arachis hypogaea
US-09-106-872A-17
Alignment Scores:
Pred. No.: 2,14e-06 Length: 335
Score: 157.50 Matches: 55
Percent Similarity: 42.04% Conservative: 48
Best Local Similarity: 22.45% Mismatches: 81
Query Match: 5.85% Indels: 61
DB: 4 Gaps: 9
US-10-728-051-3 (1-1524) x US-09-106-872A-17 (1-335)
Qy 31 TTCAGGCGCTCAATCGGAGAGCTGACAACTCGCATTCAGAGGGCGGTACATT 90
Db 122 PheGlyLysLeuPheGluValIleSfyrProAspLysLys----- 133
Qy 91 GAGACTTGGAAACCCCAACAGAGGAGTTGCAATGCGCGCGCTCGCTCTCTCGCTTA 150
Db 91 GAGACTTGGAAACCCCAACAGAGGAGTTGCAATGCGCGCGCTCGCTCTCTCGCTTA 150

Db	134	-----AsnProGlnLeuGlnAspLeuAsp-----MetMetLeuThrCysVal	147
QY	151	GTCTCCGCGCAACCCCTTCGTAGGCT---TTCTACTCCAATGCTCCCGAGAGATC	207
Db	148	GlulLeuGluGluAlaLeuMetLeuProHisPheAsnSerLysAlaMetValIleVal	167
QY	208	TTCAATCAGCAGGAAGGGGATCTTTGGTTGATATTCCTGGTTGCTTAGACACTAT	267
Db	168	ValValAsnLysGlyThrGlyAsnLeuGluLeuVal-----	179
QY	268	GAAGAGCCTCACACACAGCTCGTCTGATCTCAGTCCCAAGACCAAGACGTCCTCAA	327
Db	180	-----AlaValArgLysGluGlnGlnArgGlyArgArgGluGlu	193
QY	328	GGAGAGACCAACCAACGACGAGATAGTCACAGAGGTGCACCGTTTC-----	381
Db	194	GlulAspGluAspGluGluGluGlySerAsnArgGluValArgArgTyrThrAla	213
QY	382	-----GATCAGGTGATCTCATTCAGTTCCTCCACCGGTGCTTCTTGGCTTACAAC	435
Db	214	ArgLeuLysGluGlyAspAlaPheIleMetProAlaAlaHisProValalaIleAsnAla	233
QY	436	GACCACACACTGATGTGTGTGTTCTTCTTACTGACACCAACCAAC-----	486
Db	234	SerSerGluLeuHisLeuLeuGlyPheGlyIleAsnAlaGluAsnAsnHisArgIlePhe	253
QY	487	-----GACACACAGCTTGATCAGTTCCCGAGGAG-----	516
Db	254	LeuAlaGlyAspLysAspAsnValIleAspGlnIleGluLysGlnAlaLysAspLeuAla	273
QY	517	TTCAATTTGGTGGGACACGAGCAGAGTTCCTTAAGGTACCAACCAACCAAGCAGCAA	576
Db	274	PheProGlySerGlyGluGlnValGluLysLeuIleLysAsnGlnLysGluSerHisPhe	293
QY	577	AGCAGACGAAGAGCTTACCATATAGCCCATACAGCCCGCAAGTTCAGCCTACAGAA	636
Db	294	ValSerAlaGlnSerGlnSerProSerProSerProGluLysGlu-----	309
QY	637	GAGCGTGAATTTAGCCCTCGAGGACACGACGCGGAGAGAACGAGGACGAGCAAGA	696
Db	310	-----SerPro-----GluLysGluAspGlnGluGlu	318
QY	697	GAACAGAGGTGGA	711
Db	319	GlulAsnGlnGlyGly	323
RESULT 12			
US-07-955-905A-23			
Sequence 23, Application US/07955905A			
Patent No. 5770433			
GENERAL INFORMATION:			
APPLICANT: RECOMBINANT 47 AND 31 kD COCOA PROTEINS AND			
TITLE OF INVENTION: PRECURSOR			
NUMBER OF SEQUENCES: 28			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Floppy disk			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/07/955,905A			
FILING DATE: 21-JAN-1993			
CLASSIFICATION: 435			
INFORMATION FOR SEQ ID NO: 23:			
SEQUENCE CHARACTERISTICS:			
LENGTH: 587 amino acids			
TYPE: amino acid			
TOPOLOGY: linear			
MOLECULE TYPE: protein			
ORIGINAL SOURCE:			
ORGANISM: Gossypium hirsutum			
FEATURE:			

NAME/KEY: Protein			
LOCATION: 1..587			
OTHER INFORMATION: /note= "Vicilin from G. hirsutum"			
US-07-955-905A-23			
Alignment Scores:			
Pred. No.:	3.57e-06	Length:	587
Score:	156.50	Matches:	112
Percent Similarity:	30.60%	Conservative:	78
Best Local Similarity:	18.04%	Mismatches:	184
Query Match:	5.82%	Indels:	247
DB:	1	Gaps:	23
US-10-728-051-3 (1-1524) x US-07-955-905A-23 (1-587)			
QY	232	TTTCGGTGTGATA-----TTCCCTGGT-----	TGTCCT 258
Db	19	PheGlyLeuLeuCysSerAlaLysAspPheProGlyArgArgGlyAspAspProPro	38
QY	259	AGACACTATGAA-----GAGCTCACACACAGGTCTCGATCTCAG	300
Db	39	LysArgTyrGluAspCysArgArgCysGluTrpAspThrArgGlyGlnLysGluGln	58
QY	301	TCCCAA-----AGACCACCAACAGCTCTCAA	327
Db	59	GlnGlnCysGluGluSerCysLysSerGlnTyrClyGlyLysAspGlnGlnArgHisArg	78
QY	328	GGAGAGACCAAAAGCCACAGACAGATAGTCCACAGAGGTGCACCGTTTCGATGAG	387
Db	79	ProGluAspProGlnArgArgTyrGluGluCysGlnGlnGluCysArgGlnGlnGluGlu	98
QY	388	GGTGTATCTCATTCGAGTCCCAACCGGTGTTCTTCTGGCTCTCTACACACGACACT	447
Db	98	-----	98
QY	448	GATGTTGTGTTGTTCTTCTTACTGACACCAACAAACAGACACAGCTTGATCAGTTC	507
Db	99	-----ArgGlnArgProGlnCys	104
QY	508	CCAGAGATTCATTTGGTGGGACACGAGCAGCAGAGTTCCTTAAGGTACACACAA	567
Db	105	GlnGlnArgCys-----LleLysArgPheGluGlnGln	115
QY	568	AGCAGACAAAGCAGACGAAGCTTACCATATAGCCCATACAGCCGCAAGTACGCT	627
Db	116	GlnGlnGlnSerGlnArgGlnPheGlnGluCysGlnGlnHisCysHisGlnGlnGln	135
QY	628	AGACAAGAGACGCGTGA-----TTTAGCCCTCGAGA	660
Db	136	ArgProGluArgLysGlnGlnCysValAlaGluCysArgGluArgTyrGlnGluAsnPro	155
QY	661	CAGCAGCCGCGAGAGACGACGACGAGCAAGAGAGAAACGAGGTGGAAACATCTTC	720
Db	156	TrpArgArgGluArgGluGluGluAlaGluGluGluGluThrGluGluGlyGluGlu	175
QY	721	AGCGGCTTCACGCGC---GAGTTCTCGGAACAGCTTCCAGTTGACGACAGACAG---	774
Db	176	GlnSerHisAsnProPheHisPheHisArgArgSerPheGlnSerArgPheArgGlu	195
QY	774	-----	774
Db	196	HisGlyAsnPheArgValLeuGlnArgPheAlaSerArgHisProIleLeuArgGlyIle	215
QY	774	-----	774
Db	216	AsnGluPheArgLeuSerIleLeuGluAlaAsnProAsnThrGluValLeuProHisHis	235
QY	775	-----ATAGTGCAGAAACCTAAGAGC-----	795
Db	236	CysAspAlaGluLysIleTyrLeuValThrAsnGlyArgGlyThrLeuThrPheLeuThr	255
QY	796	-----GAGACCGAGAGTGAAGAAGAGGAGCCATTGTGACAGTGGAGGAGGC	843

256 HisGluAsnLysGluSerTyrAsnValValProGlyValValArgValProAlaGly 275
QY 844 -----CTCAGATCTTGAGCCAGATAGAAG----- 870
Db 276 SerThrValTyrLeuAlaAsnGlnAspAsnLysGluLysLeuIleAlaValLeuHis 295
QY 871 AGACGTGCCACGACGAAGAAGGAATACGATGA----- 903
Db 296 ArgProValAsnAsnProArgGlnPheGluGluPheProAlaGlySerGlnArgPro 315
QY 903 ----- 903
Db 316 GlnSerTyrLeuArgAlaPheSerArgGluIleLeuGluProAlaPheAsnThrArgSer 335
QY 904 -----GATGAATATGATACGATGAAGAGATAGAAGCGTGGCAGGGAAGCAGA 954
Db 336 GluGlnLeuAspGluLeuPheGlyGlyArgGlnSerHisArgGlnGlnGlyGlnGly 355
QY 955 GGCAGGGGGAATGGTATTGAAGACGATCTGCACCCAGTGTCTAAAGACATTTGGT 1014
Db 356 MetPheArgLysAlaSerGlnGlnGlnIleArgAlaLeuSerGlnGluAlaThrSerPro 375
QY 1015 AGAAACAGATCCCTCTAC-----ATCTACAAC 1041
Db 376 ArgGluLysSerGlyGluArgPheAlaPheAsnLeuTyrArgThrProArgTyrSer 395
QY 1042 CTCAAGCTGGT-----TCACCTAAAGTCCCAACGATCTC 1077
Db 396 AsnGlnAsnGlyArgPheTyrGluAlaCysProArgGluPheArgGlnLeuSerAspIle 415
QY 1078 AACCTTCTAATCTAGTGGCTGGACCTAGCTGCTCAATATGGAATCTCTACAGGAAT 1137
Db 416 AsnValThrVal-----SerAlaLeu-----GlnLeuAsnGlnGly 427
QY 1138 GCATTGTTCTCGCTCACTACAACCAACCAACGACACATCATATATCGATTGAGGGGA 1197
Db 428 SerIlePheValProHisTyrAsnSerLysAlaThrPheValValLeuValAsnGluGly 447
QY 1198 CGGGCTCAGTCAAGTCGTGGAC----- 1221
Db 448 AsnGlyTyrValGluMetValSerProHisLeuProArgGlnSerSerPheGluGlu 467
QY 1222 -----AGCAACGGCAACAGGTG 1239
Db 468 GluGluGlnGlnGlnGlnGlnGlnGluArgArgSerGlyGlnTyrArgLys 487
QY 1240 TAGCAGCAGGAGCTTCAAGAGGTCACGTGTGTGGTCCACAGAACTTCGCGCTCGCT 1299
Db 488 IleArgSerGlnLeuSerArgGlyAspIlePheValValProAlaAsnPheProValThr 507
QY 1300 ---GGAAAGTCCAGACGAGAACTTCGAATACGTGGCATTC----- 1338
Db 508 PheValAlaSerGlnAsnGlnAsnLeuArgMetThrGlyPheGlyLeuTyrAsnGlnAsn 527
QY 1339 ---AAGACACTCAAGCCCGCCAGTACCACTCGCGGTGAAACTCCGTCATAGAT 1395
Db 528 IleAsnProAspHisAsnGlnArgIlePhe---ValAlaGlyLysIleAsnHisValArg 546
QY 1396 AACCTGCGGAGGAGGTGGTTCAAATTCATATATGCGCTCCAAAGGAGCAGGCAAGCAG 1455
Db 547 GlnTrpAspSerGlnAlaLysGluLeuAlaPheGlyValSerSerArgLeuValAspGlu 566
QY 1456 CTTAAGAACCAACCCCTTCAAG---TTCTGCTTCCACCGTCTCAGCAGTCTCCGAGG 1512
Db 567 IlePheAsnAsnProGlnGluSerTyrPheVal-----SerArgGlnArgGlnArg 584
QY 1513 GCT 1515
Db 585 Ala 585

RESULT 13

US-07-955-905A-2

; Sequence 2, Application US/07955905A

; Patent No. 5770433
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: RECOMBINANT 47 AND 31 kD COCOA PROTEINS AND
; TITLE OF INVENTION: PRECURSOR
; NUMBER OF SEQUENCES: 28
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (SPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/955,905A
; FILING DATE: 21-JAN-1993
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 566 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-955-905A-2

Alignment Scores:
Pred. No.: 3 91e-06 Length: 566
Score: 156.00 Matches: 106
Percent Similarity: 34.82% Conservative: 74
Best Local Similarity: 20.50% Mismatches: 205
Query Match: 5.80% Indels: 132
DB: 1 Gaps: 22

US-10-728-051-3 (1-1524) x US-07-955-905A-2 (1-566)

QY 253 TGTCTAGACACTATGAAGAGCCTCACACAAGGTGTCGTCTCTCAGTCCCAAGACCA 312
Db 66 CysGluArgLysTyrLysGluGlnArgGln---GlnGluGluGluGlnArgGln 84
QY 313 CCAAGACGTCTCAAGAGAACCAAGACCAAGCCACAGCAA----- 351
Db 85 TyrGlnGlnCysGlnGlyArgCysGlnGlnGlnGlnGlnGlnGlnGln 104
QY 352 -----CGAGATAGTCACCAAGAGGTG 372
Db 105 CysGlnArgLysCysTrpGluGlnTyrLysGluGlnGluArgGlyGluHisGluAsnTyr 124
QY 373 CACCGTTTC-----GATGAGCGT-----GATCTCATT 399
Db 125 HisAsnHisLysLysAsnArgSerGluGluGluGluGlnGlnArgAsnAsnProTyr 144
QY 400 GCAGTCCCAACCGGTGTGCTTTCTGGCTC---TACAAACGACCAACGACACTGATGTTGT 456
Db 145 TyrPheProLysArgArgSerPheGlnThrArgPheArgAspGluGluGlyAsnPheLys 164
QY 457 GCTGTTTCTCTTACTGACCAACAAC-----AACGACCAACAGCTT 498
Db 165 IleGluGlnArgPheAlaGluAsnSerProPheLeuLysGlyIleAsnAspTyrArgLeu 184
QY 499 GATCAGTTC-----CCCAGGAGATTCAATTTGCTGGGAACACGCGAGCAAGAG--- 546
Db 185 AlaMetPheGluAlaAsnProAsnThrPheIleLeuProHisCysAspAlaGluAla 204
QY 547 -----TTCTTAAGGTACCAACCAACAAGC 570
Db 205 IleTyrPheValThrAsnGlyLysGlyThrIleThrPheValThrHisGluAsnLysGlu 224
QY 571 AGACAAGCAGACAGAAGA-----AGCTTACCATAT---AGCCCATACAGCCCG 615
Db 225 SerTyrAsnValGlnArgGlyThrValValSerValProAlaGlySerThrValTyrVal 244
QY 616 CAAAGTCAGCTACAGAAGAAGAGCGTGAATTTAGC-----CCTCGAGGACAG 663
Db 245 ValSerGlnAspAsnGlnGluLysLeuThrIleAlaValLeuAlaLeuProValAsnSer 264

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/955,905A
; FILING DATE: 21-JAN-1993
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 566 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Theobroma cacao
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..566
; OTHER INFORMATION: /note= "67 kD Precursor Protein"
;
US-07-955-905A-22

```

```

;
; Alignment Scores:
; Pred. No.: 3,91e-06 Length: 566
; Score: 156.00 Matches: 106
; Percent Similarity: 34.82% Conservative: 74
; Best Local Similarity: 20.50% Mismatches: 205
; Query Match: 5.80% Indels: 132
; DB: 1 Gaps: 22
;
US-10-728-051-3 (1-1524) x US-07-955-905A-22 (1-566)

```

```

;
; 253 TGTCCTAGACACTATGAAGAGCCTCACACAAAGTCTGTCATCTCAGTCCCAAGACCA 312
; 66 CysGluArgGluTyrLysGluGlnArgGln---GlnGluGluGluGlnArgGln 84
;
; 313 CCAAGACGCTCTCCAGGAGAGAGACCAAGCCAAACAGCAA--- 351
; 85 TyrGlnGlnCysGlnGlyArgCysGlnGluGlnGlnGlnGlnGlnGln 104
;
; 352 --- 372
; 105 CysGlnArgLysCysTyrGluGlnTyrLysGluGlnGluArgGlyGluHisGluAsnTyr 124
;
; 373 CACCGTTTC--- 399
; 125 HisAsnHisLysLysAsnArgSerGluGluGluGluGlnArgAsnAsnProTyr 144
;
; 400 GCAGTTCACCGGTGTTCTGCTTTCTGGCTC---TACAACGACCCAGCAGTCTGTTGTT 456
; 145 TyrPheProLysArgSerPheGlnThrArgPheArgAspGluGluGluArgLys 164
;
; 457 GCTGTTTCTTACTGACACCAACAC--- 498
; 165 IleLeuGlnArgPheAlaGluAsnSerProProLeuLysGlyIleAsnAspTyrArgLeu 184
;
; 499 GATCAGTTC--- 546
; 185 AlaMetPheGluAlaAsnProAsnThrPheIleLeuProHisCysAspAlaGluAla 204
;
; 547 --- 570
; 205 IleTyrPheValThrAsnGlyLysGlyThrIleThrPheValThrHisGluAsnLysGlu 224
;
; 571 AGACAAGACAGACGAGA--- 615
; 225 SerTyrAsnValGlnArgGlyThrValSerValProAlaGlySerThrValTyrVal 244
;
; 616 CAAAGTCAGCCTAGACAAAGAGAGCGGTGAATTTAGC--- 663
; 245 ValSerGlnAspAsnGlnGluLysLeuThrIleAlaValLeuAlaLeuProValAsnSer 264
;

```

```

;
; RESULT 14
; US-07-955-905A-22
; Sequence 22, Application US/07955905A
; Patent No. 5770433
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: RECOMBINANT 47 AND 31 kD COCOA PROTEINS AND
; TITLE OF INVENTION: PRECURSOR
; NUMBER OF SEQUENCES: 28
;

```


Search completed: April 8, 2005, 05:34:42
Job time : 72.194 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 8, 2005, 05:25:59 ; Search time 174.941 Seconds
(without alignments)
5784.401 Million cell updates/sec

Title: US-10-728-051-3
Perfect score: 2591
Sequence: 1 c9gcagcaaccgaggagaa.....ctccgagggtgtggttaa 1524

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1418010 seqs, 331997259 residues

Total number of hits satisfying chosen parameters: 2836020

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-DB=cg2 1/USPTO spoop/US10728051/runat 07042005 125048 14128/app query.fasta_1.4757
-DB=Published Applications AA -QPM=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPC=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62
-TRANS=human40.cdi -LIST=45 -DALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptto -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10728051 @CGN 1 1.663 @runat 07042005 125048 14128
-NCPU=6 -ICPU=3 -NO MWAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:
1: /cg2 6/ptodata/1/pubpaa/US07 PUBCOMB.pcp:
2: /cg2 6/ptodata/1/pubpaa/PCT NEW PUB.pcp:
3: /cg2 6/ptodata/1/pubpaa/US06 NEW PUB.pcp:
4: /cg2 6/ptodata/1/pubpaa/US06 PUBCOMB.pcp:
5: /cg2 6/ptodata/1/pubpaa/US07 NEW PUB.pcp:
6: /cg2 6/ptodata/1/pubpaa/PCTUS PUBCOMB.pcp:
7: /cg2 6/ptodata/1/pubpaa/US08 NEW PUB.pcp:
8: /cg2 6/ptodata/1/pubpaa/US08 PUBCOMB.pcp:
9: /cg2 6/ptodata/1/pubpaa/US09A PUBCOMB.pcp:
10: /cg2 6/ptodata/1/pubpaa/US09B PUBCOMB.pcp:
11: /cg2 6/ptodata/1/pubpaa/US09C PUBCOMB.pcp:
12: /cg2 6/ptodata/1/pubpaa/US09 NEW PUB.pcp:
13: /cg2 6/ptodata/1/pubpaa/US10A PUBCOMB.pcp:
14: /cg2 6/ptodata/1/pubpaa/US10B PUBCOMB.pcp:
15: /cg2 6/ptodata/1/pubpaa/US10C PUBCOMB.pcp:
16: /cg2 6/ptodata/1/pubpaa/US10D PUBCOMB.pcp:
17: /cg2 6/ptodata/1/pubpaa/US10 NEW PUB.pcp:
18: /cg2 6/ptodata/1/pubpaa/US10 NEW PUB.pcp:
19: /cg2 6/ptodata/1/pubpaa/US60 NEW PUB.pcp:
20: /cg2 6/ptodata/1/pubpaa/US60 PUBCOMB.pcp:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description
1	2673	99.3	507	15	US-10-245-871-96	Sequence 96, Appl
2	2673	99.3	507	15	US-10-253-286-96	Sequence 96, Appl
3	2673	99.3	510	14	US-10-228-806-6	Sequence 6, Appl
4	2673	99.3	510	15	US-10-100-303A-90	Sequence 90, Appl
5	2638	98.0	526	9	US-09-731-221-79	Sequence 79, Appl
6	2621	97.4	530	17	US-10-899-551-6	Sequence 6, Appl
7	1665	61.9	351	17	US-10-899-551-58	Sequence 58, Appl
8	1502	55.8	481	10	US-09-759-967-21	Sequence 21, Appl
9	1502	55.8	481	15	US-10-424-599-171702	Sequence 171702, A
10	1502	55.8	488	15	US-10-425-114-43865	Sequence 43865, A
11	1502	55.8	481	15	US-10-425-114-44047	Sequence 44047, A
12	1502	55.8	492	15	US-10-425-114-43836	Sequence 43836, A
13	1502	55.8	492	15	US-10-425-114-46294	Sequence 46294, A
14	1499.5	55.7	466	15	US-10-425-114-71959	Sequence 71959, A
15	1499.5	55.7	466	15	US-10-425-114-71973	Sequence 71973, A
16	1499.5	55.7	485	10	US-09-759-967-20	Sequence 20, Appl
17	1499.5	55.7	488	15	US-10-425-114-45645	Sequence 45645, A
18	1499.5	55.7	489	15	US-10-425-114-43983	Sequence 43983, A
19	1499.5	55.7	489	15	US-10-425-114-44676	Sequence 44676, A
20	1499.5	55.7	489	15	US-10-425-114-45612	Sequence 45612, A
21	1499.5	55.7	489	15	US-10-425-114-50117	Sequence 50117, A
22	1499.5	55.7	489	15	US-10-425-114-50207	Sequence 50207, A
23	1499.5	55.7	489	15	US-10-425-114-51412	Sequence 51412, A
24	1499.5	55.7	489	15	US-10-425-114-58429	Sequence 58429, A
25	1499.5	55.7	489	15	US-10-425-114-71902	Sequence 71902, A
26	1499.5	55.7	489	15	US-10-425-114-71909	Sequence 71909, A
27	1499.5	55.7	489	15	US-10-425-114-71957	Sequence 71957, A
28	1499.5	55.7	489	15	US-10-425-114-71958	Sequence 71958, A
29	1499.5	55.7	489	15	US-10-425-114-71967	Sequence 71967, A
30	1499.5	55.7	489	15	US-10-425-114-71995	Sequence 71995, A
31	1499.5	55.7	489	15	US-10-425-114-72287	Sequence 72287, A
32	1499.5	55.7	489	15	US-10-425-114-72288	Sequence 72288, A
33	1498	55.7	489	15	US-10-425-114-45639	Sequence 45639, A
34	1496	55.6	495	10	US-09-759-967-2	Sequence 2, Appl
35	1496	55.6	495	15	US-10-424-599-171705	Sequence 171705, A
36	1496	55.6	495	15	US-10-425-114-45626	Sequence 45626, A
37	1496	55.6	501	15	US-10-425-114-43904	Sequence 43904, A
38	1496	55.6	505	15	US-10-425-114-71930	Sequence 71930, A
39	1496	55.6	505	15	US-10-425-114-71962	Sequence 71962, A
40	1496	55.6	506	15	US-10-425-114-43901	Sequence 43901, A
41	1496	55.6	506	15	US-10-425-114-45624	Sequence 45624, A
42	1496	55.6	506	15	US-10-425-114-45680	Sequence 45680, A
43	1496	55.6	507	15	US-10-425-114-43929	Sequence 43929, A
44	1496	55.6	507	15	US-10-425-114-45765	Sequence 45765, A
45	1496	55.6	508	15	US-10-425-114-45325	Sequence 45325, A

ALIGNMENTS

RESULT 1
US-10-245-871-96
; Sequence 96, Application US/10245871
; Publication No. US20030235594A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT
; APPLICANT: XU, MINZHEN
; TITLE OF INVENTION: Ii-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: RH-2013
; CURRENT APPLICATION NUMBER: US/10/245, 871
; CURRENT FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: 10/197,000
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 905
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 96
; LENGTH: 507
; TYPE: PRT
; ORGANISM: Arachis hypogaea
US-10-245-871-96

Alignment Scores:
 Pred. No.: 1,55e-221 Length: 507
 Score: 2673.00 Matches: 507
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.33% Indels: 0
 DB: 15 Gaps: 0

US-10-728-051-3 (1-1524) x US-10-245-871-96 (1-507)

QY 1 CGGCAGCAACCGGAGGAGACGGTCCAGTTCACGGCTCAATCGCAGAGACTGAC 60
 Db 1 ArgGlnGlnProGlnGluAsnAlaCysGlnPheGlnArgLeuAsnAlaGlnArgProAsp 20
 QY 61 AATCGCATTAATCAGAGGGCGGTACATTGAGACTTGAACCCCAACACAGAGATTC 120
 Db 21 AsnArgIleGluSerGluGlyGlyTyrIleGluThrTrpAsnProAsnAsnGlnGluPhe 40
 QY 121 GAATGCGCGGCTCGCCCTCTCTCGTTAGTCTCCGCGCAACGCCCTTCGTAGGCCT 180
 Db 41 GluCysAlaGlyValAlaLeuSerArgLeuValLeuArgArgAsnAlaLeuArgArgPro 60
 QY 181 TTCTACTCCAAATGCTCCCAAGGAGATCTTCATCCAGCAAGGAGGATCTTTGGGTG 240
 Db 61 PheTyrSerAsnAlaProGlnGluIlePheIleGlnGlnGlyArgGlyTyrPheGlyLeu 80
 QY 241 ATATTCCCTGGTGTCTAGACACTATGAGAGCCTCACACAAAGTCTCGATCTCAG 300
 Db 81 IlePheProGlyCysProArgHisTyrGlnGluProHisThrGlnGlyArgArgSerGln 100
 QY 301 TCCCAAGACACCAAGACCTCCCAAGGAGAGACCAAGGAGCAACAGCAACAGATAGT 360
 Db 101 SerGlnArgProProArgArgLeuGlnGlyGluAspGlnSerGlnGlnGlnArgAspSer 120
 QY 361 CACCAGAGGTGCAACCTTTCGATGAGGTGATCTCATTCAGTTCACCGTTCACCGTGTGCT 420
 Db 121 HisGlnLysValHisArgPheAspGluGlyAspLeuIleAlaValProThrGlyValAla 140
 QY 421 TTCTGGCTTACACGACACGACACTGATGTTGTCTCTTCTTACTGACACCAAC 480
 Db 141 PheTrpLeuTyrAsnAspHisAspThrAspValValAlaValSerLeuThrAspThrAsn 160
 QY 481 AACACAGCAACACAGCTTGATCAGTCCCGCAGGAGATTCATTTGGCTGGGAACACGGAG 540
 Db 161 AsnAsnAspAsnGlnLeuAspGlnPheProArgArgPheAsnLeuAlaGlyAsnThrGlu 180
 QY 541 CAAGAGTTCTTAAGTACCAGCAACAAAGCAGCAACAAAGCAGCAACAAAGCTTACCATAT 600
 Db 181 GlnGluPheLeuArgTyrGlnGlnGlnSerArgGlnSerArgArgArgSerLeuProTyr 200
 QY 601 AGCCCATACAGCCGCAAGTCAGCTAGCAAGAGCGGTGAATTTAGCCCTCCGAGGA 660
 Db 201 SerProTyrSerProGlnSerGlnProArgGlnProArgGlnGluArgGluPheSerProArgGly 220
 QY 661 CAGCACAGCCAGCAGACAGCAGGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAATTC 720
 Db 221 GlnHisSerArgArgGluArgAlaGlyGlnGlnGluGluAsnGluGlyGlyAsnIlePhe 240
 QY 721 AGCGGCTTCAGCGGAGTTCCTGGAAACAGCCCTTCAGGTTGACGACAGACAGATAGTG 780
 Db 241 SerGlyPheThrProGluPheLeuGluGlnAlaPheGlnValAspAspArgGlnIleVal 260
 QY 781 CAACACCTTAAGGGGAGCAGCAGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840
 Db 261 GlnAsnLeuArgGlyGluThrGluSerGlnGluGluGlyAlaIleValThrValArgGly 280
 QY 841 GGCCTCAGATCTTGAAGCCAGATAGAAGAGAGCGTCCGAGCAAGAGAGAGATACGAT 900
 Db 281 GlyLeuArgIleLeuSerProAspArgIysArgArgAlaAspGluGluGluGluTyrAsp 300
 QY 901 GAAGATGATATGATACGATGAGGATAGAGCGTTCGAGGAGGAGGAGGAGGAGGAGGAGG 960

Db 301 GluAspGluTyrGluTyrAspGluGluAspArgArgArgGlySerArgGlyArg 320
 QY 961 GGGATGTTATTGAAGAGAGATCTGCACCGCAGTCTAAAAAGAACATTGTGTAGAAAC 1020
 Db 321 GlyAsnGlyIleGluGluThrIleCysThrAlaSerAlaLysLysAsnIleGlyArgAsn 340
 QY 1021 AGATCCCTTGACATCTCAACCCCTCAAGCTGTTGTTCACTCAAACTGCCAAGCATCTCAAC 1080
 Db 341 ArgSerProAspIleTyrAsnProGlnAlaGlySerLeuLysThrAlaAsnAspLeuAsn 360
 QY 1081 CTTCTAATACTAGTGTGGTGGACCTAGTCTCAATATGAAATCTCTACAGGAATGCA 1140
 Db 361 LeuLeuIleLeuArgTrpLeuGlyProSerAlaGluTyrGlyAsnLeuTyrArgAsnAla 380
 QY 1141 TTGTTTGTGCTCACTCAACACCAACGACACACAGATCATATATTCGATTGAGGGGACGG 1200
 Db 381 LeuPheValAlaHisTyrAsnThrAsnAlaHisSerIleIleTyrArgLeuArgGlyArg 400
 QY 1201 GCTCAGTGCAGTGTGGACAGCAACGGCAACAGAGTGTACGACGAGGAGGCTTCAAGAG 1260
 Db 401 AlaHisValGlnValValAspSerAsnGlyAsnArgValTyrAspGluGluLeuGlnGlu 420
 QY 1261 GGTCAAGTGTGTGGTCCACAGAACTTCGCGCTCGCTGGAAGAGTCCACAGAGGAGAAC 1320
 Db 421 GlyHisValLeuValValProGlnAsnPheAlaValAlaGlyLysSerGlnSerGluAsn 440
 QY 1321 TTCGAATACGTGGCATTCACAGACAGACTCAAGGCCAGCAGTAGCAACCTCGCGGTGAA 1380
 Db 441 PheGluTyrValAlaPheLysThrAspSerArgProSerIleAlaAsnLeuAlaGlyGlu 460
 QY 1381 AACTCGCTCATAGATAACTCGCGGAGGAGGTGTGCAAAATTCATATGCGCTCCAAAGG 1440
 Db 461 AsnSerValIleAspAsnLeuProGluGluValValAlaAsnSerTyrGlyLeuGlnArg 480
 QY 1441 GAGCAGGCAAGCAGCTTAAGAAACAAACCCCTTCAAGTTCCTCGTTCCACCGTCTCAG 1500
 Db 481 GluGlnAlaArgGlnLeuLysAsnAsnAsnProPheLysPhePheValProSerGln 500
 QY 1501 CAGTCTCCGAGGCTGTGGCT 1521
 Db 501 GlnSerProArgAlaValAla 507

RESULT 2

US-10-253-286-96
 ; Sequence 96, Application US/10253286
 ; Publication No. US2004005881A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HUMPHREYS, ROBERT
 ; APPLICANT: XU, MINZHEN
 ; TITLE OF INVENTION: II-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
 ; FILE REFERENCE: REH-2015
 ; CURRENT APPLICATION NUMBER: US/10/253,286
 ; CURRENT FILING DATE: 2003-01-13
 ; PRIOR APPLICATION NUMBER: 10/197,000
 ; PRIOR FILING DATE: 2002-07-17
 ; PRIOR APPLICATION NUMBER: 09/396,813
 ; PRIOR FILING DATE: 1999-09-14
 ; NUMBER OF SEQ ID NOS: 905
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 96
 ; LENGTH: 507
 ; TYPE: PRT
 ; ORGANISM: Arachis hypogaea
 ; US-10-253-286-96

Alignment Scores:
 Pred. No.: 1,55e-221 Length: 507
 Score: 2673.00 Matches: 507
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.33% Indels: 0
 DB: 15 Gaps: 0

US-10-728-051-3 (1-1524) x US-10-253-286-96 (1-507)

```
QY 1 CGGACGACCGGAGGAGACGGCTGCCAGTTCAGCGCCTCAATGCGCAGACCTTGAC 60
Db 1 ArgGlnGlnProGluGluAsnAlaCysGlnPheGlnArgLeuAsnAlaGlnArgProAsp 20
QY 61 AATCGCATTCAATCAGAGGCGGTACATTGAGCTTGGAAACCCCAACACCGAGGATTC 120
Db 21 AsnArgIleGluSerGluGlyGlyTyrlleGluThrTrpAsnProAsnAsnGlnGluPhe 40
QY 121 GAATGCGCGCGCTGCGCCTCTCTCGCTTAGTCTCCGCGCGCAACGCGCTTCGTAGGCCT 180
Db 41 GluCysAlaGlyValAlaLeuSerArgLeuValLeuArgAsnAlaLeuArgPro 60
QY 181 TTCTACTCCAATGCTCCCGAGGATCTTCATCCAGCAAGAAAGGGGATCTTTGGGTTG 240
Db 61 PheTyrSerAsnAlaProGlnGluIlePheIleGlnGlnGlyArgGlyTyrPheGlyLeu 80
QY 241 ATATTCCTCGTGTCTAGACACTATGAAGAGCTCACACAGAGTTCGTCATCTCAG 300
Db 81 IlePheProGlyCysProArgHisTyrGluGluProHisThrGlnGlyArgGserGln 100
QY 301 TCCCAAGACACCAAGACGCTCTCAAGGAGAGACCAAGCCAAAGCAACACGAGATAGT 360
Db 101 SerGlnArgProProArgArgLeuGlnGlyGluAspGlnSerGlnGlnArgAspSer 120
QY 361 CACCAGAGGTGACCGTTCGATGAGGGTATCTCATTCAGATTCACCGGTGTTGCT 420
Db 121 HisGlnLysValHisArgPheAspGluGlyAspLeuIleAlaValProThrGlyValAla 140
QY 421 TTCTGGCTCTACACGACCGACACTGATGTTGTTGCTGTTCTCTTCTTACTGACCCAAC 480
Db 141 PheTrpLeuTyrAsnAspHisAspThrAspValValAlaValSerLeuThrAspThrAsn 160
QY 481 AACCAACCAACACAGCTTGATCAGTCCCGCAGGAGATTCAATTGGCTGGGAAACACCGAG 540
Db 161 AsnAsnAspAsnGlnLeuAspGlnPheProArgArgPheAsnLeuAlaGlyAsnThrGlu 180
QY 541 CAAGAGTCTTAAGGTACCGACCAACAAAGCAGACAAAGCAGACGAAAGCTTACCATAT 600
Db 181 GlnGluPheLeuArgTyrGlnGlnSerArgGlnSerArgArgArgSerLeuProTyr 200
QY 601 AGCCCATACGCCGCGCAAGTACGCTTAGACAGAGAGCGTGNATTTAGCCCTCGAGGA 660
Db 201 SerProTyrSerProGlnSerGlnProArgGlnGluArgGluPheSerProArgGly 220
QY 661 CAGCACAGCGCAGACAGACGAGGACAAAGAAAGAAACGAAGTGGAAACATCTTC 720
Db 221 GlnHisSerArgArgGluArgAlaGlyGlnGluGluAsnGluGlyGlyAsnIlePhe 240
QY 721 AGCGGCTTCAGCGCGAGTTCCTGGAAACAGCCCTTCAGGTTGACGACAGACAGATAGTG 780
Db 241 SerGlyPheThrProGluPheLeuGluGlnAlaPheGlnValAspAspArgGlnIleVal 260
QY 781 CAACACCTTAGCGCGAGACCGAGAGTGAAGAGAGGAGGCCATTGTGACAGTGAGGGA 840
Db 261 GlnAsnLeuArgGlyGluThrGluSerGluGluGluGlyAlaIleValThrValArgGly 280
QY 841 GGCTCTCAGAATCTTGACCCAGATAGAAAGAGAGCGTCCCGACGAAGAGAGGAATACGAT 900
Db 281 GlyLeuArgIleLeuSerProAspArgLysArgArgAlaAspGluGluGluGluTyrAsp 300
QY 901 GAAGATGAATATGAATACGATGAAGAGAGATAGAAGCGTGGCAGGGAAGCAGAGGCGAG 960
Db 301 GluAspGluTyrGluTyrAspGluGluAspArgArgArgGlyArgGlySerArgGlyArg 320
QY 961 GGGATCGTATTGAAGAGACGATCTGCACCGCAGTCTCTAAAGAACATTTGGTAGAAC 1020
Db 321 GlyAsnGlyIleGluGluThrIleCysThrAlaSerAlaLysLysAsnIleGlyArgAsn 340
QY 1021 AGATCCCTGACATCTACACCTCAAGCTGGTTCACTCAAAAGTCCCAACGATCTCAAC 1080
Db 341 ArgSerProAspIleTyrAsnProGlnAlaGlySerLeuLysThrAlaAsnAspLeuAsn 360
```

```
QY 1081 CTTCTAATACTTAGTGGCTTGACCTAGTCTGAATATGGAATCTCTACAGGAATGCA 1140
Db 361 LeuLeuIleLeuArgTrpLeuGlyProSerAlaGluTyrGlyAsnLeuTyrArgAsnAla 380
QY 1141 TTCTTTGTGCTCCTACCAACCAACGACGACACGATCATATATCGATTGAGGGACGG 1200
Db 381 LeuPheValAlaHisTyrAsnThrAsnAlaHisSerIleIleTyrArgLeuArgGlyArg 400
QY 1201 GCTCAGCTGCAAGTCTGTGACAGCAACGCGACAGAGTGATACGACGAGGAGCTTCAAGAG 1260
Db 401 AlaHisValGlnValValAspSerAsnGlyAsnArgValTyrAspGluGluLeuGlnGlu 420
QY 1261 GGTCAAGTCTGTGTGTCGACAGAACTTCGCGCTCGTGGAAAGTCCCGAGGAGAAC 1320
Db 421 GlyHisValLeuValValProGlnAsnPheAlaValAlaGlyLysSerGlnSerGluAsn 440
QY 1321 TTCGAATACGTGGCATTCAAGACAGACTCAAGGCCAGCATAGCAACCTCGCGGTGAA 1380
Db 441 PheGluTyrValAlaPheLysThrAspSerArgProSerIleAlaAsnLeuAlaGlyGlu 460
QY 1381 AACTCCGTCATAGATAACCTGCGGAGAGGTGGTTGCAAAATTCATATGGCCTCCAAAGG 1440
Db 461 AsnSerValIleAspAsnLeuProGluValValAlaAsnSerTyrGlyLeuGlnArg 480
QY 1441 GACGAGCGAAGGAGCGCTTAAGAACAACAACCCCTTCAAGTCTTCGTTCCACCGTCTCAG 1500
Db 481 GluGlnAlaArgGlnLeuLysAsnAsnProPheLysPheValProProSerGln 500
QY 1501 CAGTCTCCGAGGGCTGTGGCT 1521
Db 501 GlnSerProArgAlaValAla 507
```

RESULT 3

```
US-10-228-806-6
; Sequence 6, Application US/10228806
; Publication No. US20030049237A1
; GENERAL INFORMATION:
; APPLICANT: Bannon, et al.
; TITLE OF INVENTION: Methods and Reagents for Decreasing Clinical Reactions
; TITLE OF INVENTION: to Allergy
; FILE REFERENCE: 2002834-0043
; CURRENT APPLICATION NUMBER: US/10/228,806
; CURRENT FILING DATE: 2002-08-26
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Arachis hypogaea
US-10-228-806-6
```

```
Alignment Scores:
Pred. No.: 1 55e-221 Length: 510
Score: 2673.00 Matches: 507
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.33% Indels: 0
DB: 14 Gaps: 0
```

US-10-728-051-3 (1-1524) x US-10-228-806-6 (1-510)

```
QY 1 CGGACGACCGGAGGAGACGGCTGCCAGTTCAGCGCCTCAATGCGCAGACCTTGAC 60
Db 4 ArgGlnGlnProGluGluAsnAlaCysGlnPheGlnArgLeuAsnAlaGlnArgProAsp 23
QY 61 AATCGCATTCAATCAGAGGCGGTTCATTGAGACTTGGAAACCCCAACACCGAGGATTC 120
Db 24 AsnArgIleGluSerGluGlyGlyTyrIleGluThrTrpAsnProAsnAsnGlnGluPhe 43
QY 121 GAATGCGCGCGCTGCGCCTCTCTCGTTAGTCTCCCGCGCAACGCGCTTCGTAGGCCT 180
Db 44 GluCysAlaGlyValAlaLeuSerArgLeuValLeuArgArgAsnAlaLeuArgArgPro 63
```

```
QY 181 TTCTACTCCAATGCTCCCGAGAGATCTTCATCCAGCAAGAAAGGGATACATTTGGGTG 240
Db 64 PheTyrSerAsnAlaProGlnGluilePheIleGlnGlnGlyArgGlyTyrPheGlyLeu 83
QY 241 ATATTCCCTGGTGTCTAGACACTATGAAGAGCCTCACACAAAGTGTGTCATCTCAG 300
Db 84 IlePheProGlyCysProArgHisTyrGluGluProHisThrGlnGlyArgArgSerGln 103
QY 301 TCCCAAGACCAACAGAGCTCTCCAGGAGAGAGACAAAGCCAAACAGCAACGAGATAGT 360
Db 104 SerGlnArgProArgArgLeuGlnGlnGlyLysAspGlnSerGlnGlnArgAspSer 123
QY 361 CACCAGAAGTGCACCGTTCGATGAGGTGATCTCATTCAGTCTCCACCGGTGTGTCT 420
Db 124 HisGlnLysValHisArgPheAspGluGlyAspLeuIleAlaValProThrGlyValAla 143
QY 421 TTCTGGCTCTACAAACACGACGACTGATGTTGTTGCTGTTCTTCTTACTGACACCAAC 480
Db 144 PheTrpLeuTyrAsnAspHisAspThrAspValValAlaValSerLeuThrAspThrAsn 163
QY 481 AACAAACGACACAGCTTGATCAGTTCCCGAGGAGATTCAATTTGCTGGCAACCGAG 540
Db 164 AsnAsnAspGlnGlnLeuAspGlnPheProArgPheAsnLeuAlaGlyAsnThrGlu 183
QY 541 CAAGAGTCTTAAAGTACCAAGCAACAAAGCAGACAAAGCAGACGAAAGCTTACCATAT 600
Db 184 GlnGluPheLeuArgTyrGlnGlnSerArgGlnSerArgArgSerLeuProTyr 203
QY 601 AGCCCATACAGCCGCGAAGTACGCTAGCAGACAGAGAGCGTGAATTTAGCCCTCAGGA 660
Db 204 SerProTyrSerProGlnSerGlnProArgGlnGluGluArgGluPheSerProArgGly 223
QY 661 CAGCAGACCGCAGAGAAACAGCAGACGACAGAAAGAAACAAAGTGGAAACATCTTC 720
Db 224 GlnHisSerArgArgGluArgAlaGlyGlnGluGluAsnGlnGlyAsnIlePhe 243
QY 721 AGCGGTTCACGCGGAGTCTCTGGAACAAAGCCTTCAGGTTCCACACAGACAGATAGTG 780
Db 244 SerGlyPheThrProGluPheLeuGluGlnAlaPheGlnValAspArgGlnIleVal 263
QY 781 CAAACCTAGAGCGCAGACCGAGAGTGAAGAGAGGAGCCATTCTGACAGTGGGGA 840
Db 264 GlnAsnLeuArgGlyGluThrGluSerGluGluGluGlyAlaIleValThrValArgGly 283
QY 841 GGCTCAGAACTTTGAGCCGACATAGAAAGAGACGTGCCGACCAAGAGGAATACGAT 900
Db 284 GlyLeuArgIleLeuSerProAspArgLysArgAlaAspGluGluGluTyrAsp 303
QY 901 GAAGATGAATATGAATACGATGAAGAGATAGAAGCGGTGGCAGGGGAAGCAGAGCAGG 960
Db 304 GluAspGluTyrGluTyrAspGluGluAspArgArgArgGlySerArgGlyArg 323
QY 961 GGAATGGTATTGAAGAGAGTCTCCACCGCAAGTGTCTAAAGAACATTGTGTAAGAAC 1020
Db 324 GlyAsnGlyIleGluGluThrIleCysThrAlaSerAlaLysLysAsnIleGlyArgAsn 343
QY 1021 AGATCCCTGACATCTACAAACCTCTCAAGCTGGTTCATCTCAAACTGCCAACGATCTCAAC 1080
Db 344 ArgSerProAspIleTyrAsnProGlnAlaGlySerLeuLysThrAlaAsnAspLeuAsn 363
QY 1081 CTTCTTAATACTTGGTGGCTTGGACCTAGTGTGAATATGAAATCTCTACAGGAATGCA 1140
Db 364 LeuLeuIleLeuArgTrpLeuGlyProSerAlaGluTyrGlyAsnLeuTyrArgAsnAla 383
QY 1141 TTGTTTGTCCCTCAGTACACCAACGACGACAGCATCATATATCGATTGAGGGACGG 1200
Db 384 LeuPheValAlaHisTyrAsnThrAsnAlaHisSerIleIleTyrArgLeuArgGlyArg 403
QY 1201 GCTCAGCTGCAAGTCTGTGACAGCAACGCAACAGAGTGTACACGAGGAGCTTCAAGAG 1260
Db 404 AlaHisValGlnValAspSerAsnGlyAsnArgValTyrAspGluGluLeuGlnGlu 423
```

```
QY 1261 GGTCACTGCTGTGTGTGTCGCCACAGAACTTCGCCGTGCTGGAAGTCCAGAGCGAGAAC 1320
Db 424 GlyHisValLeuValValProGlnAsnPheAlaValAlaGlyLysSerGlnSerGluAsn 443
QY 1321 TTCGAATAGCTGGCATTCAAGACAGACTCAAGGCCAGCATAGCCAACTCCCGGTGAA 1380
Db 444 PheGluTyrValAlaPheLysThrAspSerArgProSerIleAlaAsnLeuAlaGlyGlu 463
QY 1381 AACTCGTCTATAGATAAACCCTCGCGAGGAGGTGTGCAAAATTCATATATGCGCTCCAAAGG 1440
Db 464 AsnSerValIleAspAsnLeuProGluGluValValAlaAsnSerTyrGlyLeuGlnArg 483
QY 1441 GAGCAGGCAAGCAGCTTAAAGAACAAACACCCCTTCAAGTCTTCTGTTCCACCGTCTCAG 1500
Db 484 GluGlnAlaArgGlnLeuLysAsnAsnProPheLysPhePheValProProSerGln 503
QY 1501 CAGTCTCCGAGGGCTGTGGCT 1521
Db 504 GlnSerProArgAlaValAla 510
RESULT 4
; Sequence 90, Application US/10100303A
; Publication No. US20030202980A1
; GENERAL INFORMATION:
; APPLICANT: Caplan, et al.
; TITLE OF INVENTION: Methods and Reagents for Decreasing Clinical Reaction
; TITLE OF INVENTION: to Allergy
; FILE REFERENCE: 2002834-0166
; CURRENT APPLICATION NUMBER: US/10/100,303A
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 90
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Arachis hypogaea, Prot/Nucleo Ara h 3
US-10-100-303A-90
Alignment Scores:
Pred. No.: 1-55e-221 Length: 510
Score: 2673.00 Matches: 507
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.33% Indels: 0
DB: 15 Gaps: 0
US-10-728-051-3 (1-1524) x US-10-100-303A-90 (1-510)
QY 1 CGGCAGCAACCGGAGAGAACCGGTGCCAGTTTCCAGCGCTCAATGCGCAGAGACCTGAC 60
Db 4 ArgGlnGlnProGluGluAsnAlaCysGlnPheGlnArgLeuAsnAlaGlnArgProAsp 23
QY 61 AATCCCATTAATCAGAGCGCGTTACATTGAGACTTGGAAACCCCAACACAGAGATTC 120
Db 24 AsnArgIleGluSerGluGlyGlyTyrIleGluThrTrpAsnProAsnAsnGlnGluPhe 43
QY 121 GAATCGCGCGCGTCCCGCTCTCTCGCTAGTCTCCGCGCAACGCCCTTCGTAGCGCT 180
Db 44 GluCysAlaGlyValAlaLeuSerArgLeuValLeuArgAsnAlaLeuArgPro 63
QY 181 TTCTACTCTCAATGCTCCCGAGAGATCTTCATCCAGCAAGGAGGGATACATTTGGGTG 240
Db 64 PheTyrSerAsnAlaProGlnGluilePheIleGlnGlnGlyArgGlyTyrPheGlyLeu 83
QY 241 ATATTCCCTGGTGTCTAGACACTATGAAGAGCCTCACACAAAGTGTGTCATCTCAG 300
Db 84 IlePheProGlyCysProArgHisTyrGluGluProHisThrGlnGlyArgArgSerGln 103
QY 301 TCCCAAGACCAACAGAGCTCTCCAGGAGAGAGACAAAGCCAAACAGCAACGAGATAGT 360
Db 104 SerGlnArgProArgArgLeuGlnGlnGlyLysAspGlnSerGlnGlnArgAspSer 123
```

```
QY 361 CACCAGAGGTGCACCGTTTCGATGAGGTGATCTCATTCAGTTCACCGGTGTGCT 420
Db 124 HieGlnlyValHisArgPheAspGluGlyAspLeuileAlaValProThrGlyValAla 143
QY 421 TTCTGGCTCTACACGACGACGACACTGATGTTGTCTGTTCTCTTACTGACACCAAC 480
Db 144 PheTrpLeuTyzAsnAspHisAspThrAspValValAlaValSerLeuThrAspThrAsn 163
QY 481 AACAAACGACACCGCTTGATGATTCCTCCAGGAGATTCAATTTGGCTGGAAACACGGAG 540
Db 164 AsnAsnAspAsnGlnLeuAspGlnPheProArgPheAsnLeuAlaGlyAsnThrGlu 183
QY 541 CAAGAGTCTTAAGTACCAGCAACAAGCAGACAAAGCAGACGACGACGACGACGACGACG 600
Db 184 GlnGluPheLeuArgTyzGlnGlnGlnSerArgGlnSerArgArgSerLeuProTyz 203
QY 601 AGCCCATACAGCCGCGAAAGTACGCTTAGACAGAGAGCGTGAATTTAGCCCTCCAGGA 660
Db 204 SerProTyzSerProGlnSerGlnProArgGlnGluArgGluPheSerProArgGly 223
QY 661 CAGCAGCCGCGAGAGACGAGCAGGACGAGCAAGAAAGAAAGAGAGGAGGAGGAGGAGG 840
Db 224 GlnHisSerArgArgGluArgAlaGlyGlnGluGluAsnGluGlyGlyAsnIlePhe 243
QY 721 AGCGCTTCACGCGGAGTCTCTGGAACAAGCCCTCCAGGTTGACGACGACAGATAGTG 780
Db 244 SerGlyPheThrProGluPheLeuGluGlnAlaPheGlnValAspAspArgGlnIleVal 263
QY 781 CAAACCTTAGAGCGAGACCGAGAGTGAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 840
Db 264 GlnAsnLeuArgGlyGluThrGluSerGluGluGluGlyAlaIleValThrValArgGly 283
QY 841 GGCTTCAGAAATCTTGACCCAGATAGAAAGAGAGCGTCCGACGAAAGAGGAAATACGAT 900
Db 284 GlyLeuArgIleLeuSerProAspArgLysArgAlaAspGluGluGluTyzAsp 303
QY 901 GAAGATGAATATGAATACGATGAAGAGATAGAAGCGTGGCGAGGAGGAGGAGGAGGAGG 960
Db 304 GluAspGluTyzGluTyzAspGluGluAspArgArgGlyArgGlySerArgGlyArg 323
QY 961 GCGAATGGTATGAAGAGAGCGATCGCACCGCAAGTCTAAAGACACATTGCTAGAAAC 1020
Db 324 GlyAsnGlyIleGluGluThrIleCysThrAlaSerAlaLysLysAsnIleGlyArgAsn 343
QY 1021 AGATCCCTGACATCTACAAACCTCAAGCTGGTTCACTAAAACATGCAACGATCTCAAC 1080
Db 344 ArgSerProAspIleTyzAsnProGlnAlaGlySerLeuLysThrAlaAsnAspLeuAsn 363
QY 1081 CTTCTAATACTTAGTGGCTTGACCTAGTGTGCTGAATATGGAATCTCTACAGAAATGCA 1140
Db 364 LeuLeuileLeuArgTrpLeuGlyProSerAlaGluTyzGlyAsnLeuTyzArgAsnAla 383
QY 1141 TTCTTTGTCCTCACTACACACCAAGCAGCAGCATCATATATCGATTGAGGAGCGG 1200
Db 384 LeuPheValAlaHisTyzAsnThrAsnAlaHisSerIleIleTyzArgLeuArgGlyArg 403
QY 1201 GCTCAGCTGAAGTCGTGGACAGCAACGCAACAGAGTGATACACGAGGAGCTTCAAGAG 1260
Db 404 AlaHisValGlnValValAspSerAsnGlyAsnArgValTyzAspGluGluLeuGlnGlu 423
QY 1261 GGTCACTGCTTGTGTGTCACAGAACTTCGCCGTGCTCGAAAGTCCAGAGCGAGAAC 1320
Db 424 GlyHisValLeuValProGlnAsnPheAlaValAlaGlyLysSerGlnSerGluAsn 443
QY 1321 TTGGAATACGTGCTTCAGACAGACTCAAGCCCGAGCATAGCCACCTCGCGGTGAA 1380
Db 444 PheGluTyzValAlaPheLysThrAspSerArgProSerIleAlaAsnLeuAlaGlyGlu 463
QY 1381 AACTCCCTCATAGATAACCTGCCGAGGAGGTGTTGCAAAATCATATGTCCTCCAAAGG 1440
Db 464 AsnSerValIleAspAsnLeuProGluGluValValAlaAsnSerTyzGlyLeuGlnArg 483
QY 1441 GAGCAGGCAAGGCAGCTTAAGAAACAACACCCCTTCAAGTTCTTCGTTCCACCGTCTCAG 1500
```

```
Db 484 GluGlnAlaArgGlnLeuLysAsnAsnProPheLysPheValProPserGln 503
QY 1501 CAGTCTCCGAGGCGTGTGGCT 1521
Db 504 GlnSerProArgAlaValAla 510

RESULT 5
US-09-731-221-79
; Sequence 79, Application US/09731221
; Patent No. US20020018778A1
; GENERAL INFORMATION:
; APPLICANT: Caplan, Michael
; TITLE OF INVENTION: Passive Desensitization
; FILE REFERENCE: 2002834-0103
; CURRENT APPLICATION NUMBER: US/09/731,221
; CURRENT FILING DATE: 2001-12-06
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 79
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Arachis
; OTHER INFORMATION: Hypogaea
; US-09-731-221-79

Alignment Scores:
Pred. No.: 1,64e-218 Length: 526
Score: 2638.00 Matches: 502
Percent Similarity: 99.01% Conservative: 0
Best Local Similarity: 99.01% Mismatches: 5
Query Match: 98.03% Indels: 0
DB: 9 Gaps: 0

US-10-728-051-3 (1-1524) x US-09-731-221-79 (1-526)
QY 1 CGGCACCAACCGGAGGAGAACGCGTCCAGTTCACGCGCTCAATGCGCAGACCTGAC 60
Db 5 ArgGlnGlnProGluGluAsnAlaCysGlnPheGlnArgLeuAsnAlaGlnArgProAsp 24
QY 61 AATCGCATTAATCAGAGGCGGTATCATTTGAGACTTGGAAACCCCAACCAACAGGAGTTC 120
Db 25 AsnArgIleGluSerGluGlyGlyTyzIleGluThrTrpAsnAlaAsnGlnGluPhe 44
QY 121 GAATGCGCGGCGTCCGCTCTCTCGCTTAGTCTCGCGCAACCGCTTCGTAGGCTT 180
Db 45 GluCysAlaGlyValAlaLeuSerArgLeuValLeuArgAsnAlaLeuArgArgPro 64
QY 181 TTCTACTCCAATGCTCCCGAGAGATCTTCATCCAGCAAGAAAGGAGTACTTTGGGTTG 240
Db 65 PheTyzSerAsnAlaProGlnGluIlePheIleGlnGlnGlyArgGlyTyzPheGlyLeu 84
QY 241 ATATTCCTCGTGTGCTTAGACACTATGAAGAGCTTCACACAAAGTCTCGATCTCAG 300
Db 85 IlePheProGlyCysProArgHisTyzGluProHisThrGlnGlyArgArgSerGln 104
QY 301 TCCCAAGACCAACAGAGCTCTCCAGGAGAGACCAAGCAACAGCAACGAGATAGT 360
Db 105 SerGlnArgProProArgArgLeuGlnGlyGluAspGlnSerGlnGlnArgAspSer 124
QY 361 CACCAAGAGTGCACCGTTTCGATGAGGAGTATCTCATTCAGTTCCTCCCGGTGTGCT 420
Db 125 HisGlnLysValHisArgPheAspGluGlyAspLeuIleAlaValProThrGlyValAla 144
QY 421 TTCTGCTCTTACAAACGACCAACGACACTGATGTTGTGTTGCTGTTCTTACTGACCAAC 480
Db 145 PheTrpLeuTyzAsnAspHisAspThrAspValValAlaValSerLeuThrAspThrAsn 164
QY 481 AACAAACGACACCGCTTGATGATTCCTCCAGGAGATTCAATTTGGCTGGAAACACGGAG 540
Db 165 AsnAsnAspAsnGlnLeuAspGlnPheProArgPheAsnLeuAlaGlyAsnThrGlu 184
```

541 CAAGAGTTCTTAAGTACAGCAACAAAGCAGACAAAGCAGACGAAAGAGCTTACCATAT 600
185 GlnGluPheLeuArgTyrGlnGlnGlnSerArgGlnSerArgArgSerLeuProTyr 204
601 AGCCCATACAGCCCGCAAGTACCTAGCAGAGAGAGCGTGAATTTAGCCCTCGAGCA 660
205 SerProTyrSerProGlnSerGlnProArgGlnGluArgGluPheSerProArgGly 224
661 CACACAGCCGACAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 720
225 GlnHisSerArgArgGluArgAlaGlyGlnGlnGluGluGluGluGlyGlyAsnIlePhe 244
721 AGCGGTTTCAAGCGGAGTCTTGGAAACAAGCCCTCCAGGTTGACACACAGACATAGTG 780
245 SerGlyPheThrProGluAlaLeuPheGlnAlaPheGlnValAspArgGlnIleVal 264
781 CAAACCTTAGAGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 840
265 GlnAsnLeuArgGlyGluThrGluSerGluGluGluGluGlyAlaIleValThrValArgGly 284
841 GGCTTCAGATCTTGAAGCCAGATAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
285 GlyLeuArgAlaLeuSerProAspArgLysArgArgAlaAspGluGluGluTyrAsp 304
901 GAAGATGAATATGAATACGATGAAGAGGATAGAAGCGGTGGCAGGGAAGCAGAGCAGG 960
305 GluAspGluTyrAlaTyrAspGluGluAspArgArgGlyArgGlySerArgGlyArg 324
961 GGAATGTTTGAAGAGAGCTCTCCAGCGCAAGTCTTAAAGAACATTTGGTAGAAGC 1020
325 GlyAsnGlyIleGluGluThrIleCysThrAlaSerAlaLysLysAsnIleGlyArgAsn 344
1021 AGATCCCTCAGATCTTCAACACCTCAAGCTGGTTTCACTCAAACTGCCAAGCATCTCAAC 1080
345 ArgSerProAspIleTyrAsnProGlnAlaGlySerLeuLysThrAlaAsnAspLeuAsn 364
1081 CTTCTAATCTTAAAGTGGCTGACCTAGTGTGTAATATGGAATCTTACAGGAATGCA 1140
365 LeuLeuIleLeuArgTrpLeuGlyProSerAlaGluTyrGlyAsnLeuTyrArgAsnAla 384
1141 TTCTTTGTCTGCTCAGTACACCAACCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1200
385 LeuPheValAlaHisTyrAsnThrAsnAlaHisSerIleIleTyrArgLeuArgGlyArg 404
1201 GCTCAGCTGCAAGTCTGGACAGCAACGCGCAACAGAGTGTACAGCAGAGCTTCAAGAG 1260
405 AlaHisValGlnValValAspSerAsnGlyAsnArgValTyrAspGluGluLeuGlnGlu 424
1261 GGTCACTGCTTGTGTGTCACAGAACTTCCGCTGCTGGAAGTCCAGAGCAGAGAC 1320
425 GlyHisValLeuValValProGlnAsnPheAlaValAlaGlyLysSerGlnSerGluAsn 444
1321 TTCGAATAGCTGCTTCAAGCAGACTCAAGCCAGCAGATAGCCAACTCCGCGGTGAA 1380
445 PheGluTyrValAlaPheLysThrAspSerArgProSerIleAlaAsnLeuAlaGlyGlu 464
1381 AACTCCGCTCATAGATAACCTGCCGAGGAGGTGGTTGCCAAATTCATATGCTCCCAAGG 1440
465 AsnSerValIleAspAsnLeuProGluGluValValAlaAsnSerTyrGlyLeuGlnArg 484
1441 GACAGCGCAAGGAGCTTAAAGAACAAACCCCTTCAAGTCTTCTCCAGCTCTCAG 1500
485 GluGlnAlaArgGlnLeuLysAsnAsnAsnProPheLysPheValProProSerGln 504
1501 CAGTCTCCGAGGGCTGTGGCT 1521
505 GlnSerProArgAlaValAla 511

RESULT 6
US-10-899-551-6
: Sequence 6, Application US/10899551
: Publication No. US20050063994A1

GENERAL INFORMATION:
; APPLICANT: Caplan, Michael J.
; APPLICANT: Burks, A. Wesley
; APPLICANT: Sampson, Hugh A.
; APPLICANT: Howard, Sosin B.
; APPLICANT: Bottomly, Kim H.
; TITLE OF INVENTION: Methods and Reagents for Decreasing Clinical Reaction to Allergy
; FILE REFERENCE: 2002834-0233
; CURRENT APPLICATION NUMBER: US/10/899,551
; CURRENT FILING DATE: 2004-07-26
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 530
; TYPE: PRT
; ORGANISM: species Arachis hypogea
US-10-899-551-6

Alignment Scores:
Pred. No.: 4,83e-217 Length: 530
Score: 2621.00 Matches: 500
Percent Similarity: 98.62% Conservative: 0
Best Local Similarity: 98.62% Mismatches: 7
Query Match: 97.40% Indels: 0
DB: 17 Gaps: 0

US-10-728-051-3 (1-1524) x US-10-899-551-6 (1-530)

QY 1 CGGCAGCAACCGAGAGAGAACCGCTGCCAGCTCAATGCGCAGAGACCTGAC 60
Db 24 ArgGlnGlnProGluGluAsnAlaCysGlnPheGlnArgLeuAsnAlaGlnArgProAsp 43
QY 61 AATCGCATTTGAATCAGAGCGGCTTACATGTAGACTTGGAACTTGGAACTTGGAGTTTC 120
Db 44 AsnArgIleGluSerGluGlyGlyTyrIleGluThrTrpAsnProAsnAsnGlnGluPhe 63
QY 121 GAATGGCGCGGCTCGCCCTCTCTCGCTTAGTCTCCGCCGCAAGCCCTTCGTAGGCTT 180
Db 64 GluCysAlaGlyValAlaLeuSerArgLeuValLeuArgAsnAlaLeuArgArgPro 83
QY 181 TTCTACTCTCAATGCTCCCGAGGAGATCTTCATCCAGCAAGAGGAGTACTTTGGGTG 240
Db 84 PheTyrSerAsnAlaProGlnGluIlePheIleGlnGlnGlyArgGlyTyrPheGlyLeu 103
QY 241 ATATTCCCTGTTGTCTTAGACACTATGAAGAGCTTCACACAAAGGTCGTGATCTCAG 300
Db 104 IlePheProGlyCysProSerThrTyrGluGluProAlaGlnGlnGlyArgArgSerGln 123
QY 301 TCCCAAGACCAACCAAGAGCTCTCCAGGAGAGAGCAAGCAAGCAAGCAAGCAAGATGT 360
Db 124 SerGlnArgProProIleArgLeuGlnGlyGluAspGlnSerGlnGlnGlnArgAspSer 143
QY 361 CACCAAGAGTGCACCGTTTCGATCAGGCTGATCTCATTGACAGTTCACACCGGTGTGCT 420
Db 144 HisGlnLysValHisArgPheAspGluGlyAspLeuIleAlaValProThrGlyValAla 163
QY 421 TTCTGCTCTTACACGACCCACGACACTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 480
Db 164 PheTrpLeuTyrAsnAspHisAspThrAspValValAlaValSerLeuThrAspThrAsn 183
QY 481 ACACAGCAACACAGCTGTGATCAGTCTCCAGAGATTCATTTGGCTGGGACACGGAG 540
Db 184 AsnAsnAspAsnGlnLeuAspGlnPheProArgArgPheAsnLeuAlaGlyAsnHisGlu 203
QY 541 CAAGAGTTCCTTAAGGTACCAGCAACAAAGCAGCAAAAGCAGACGAAAGAGCTTACCATAT 600
Db 204 GlnGluPheLeuArgTyrGlnGlnGlnSerArgGlnSerArgArgSerLeuProTyr 223
QY 601 AGCCCATACAGCCCGCAAGTACAGCTCAGCTACAAAGAGAGCGGTGAATTTAGCCCTCGAG 660
Db 224 SerProTyrSerProGlnSerGlnProArgGlnGluArgGluPheSerProArgGly 243
QY 661 CAGCAGACCGCAGAGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 720

```
Db 244 GlnHisSerArgArgGluArgAlaGlyGlnGluGluGlnGluGlyGlyAsnIlePhe 263
Qy 721 AGCGGCTTACGCGCGAGTTCTCTGGAAACAAGCCCTTCCAGGTTGACGACAGACAGATAGTG 780
Db 264 SerGlyPheThrProGluPheLeuGluGlnAlaPheGlnValAspAspArgGlnIleVal 283
Qy 781 CAAACCTAAGACGCGAGACCGAGAGTGAAGAGAGGAGGACCATTTGACAGTGAGGGGA 840
Db 284 GlnAsnLeuArgGlyGluAsnGluSerGluGluGlyAlaIleValThrValArgGly 303
Qy 841 GGCCTCAGAACTTTGAGCCAGATAGAAAGAGAGCGTCCGACGAAGAAGAGGAAATAGAT 900
Db 304 GlyLeuArgIleLeuSerProAspArgArgAlaAspGluGluGluGluGluArgP 323
Qy 901 GAAGATCAATATGATACGATGAAGAGGATAGAGGCGTGCAGGGGAAGCAGAGGCGAGG 960
Db 324 GluAspGluTyrGluTyrAspGluAspArgArgGlyArgGlySerArgGlyArg 343
Qy 961 GGGATGCTATTGAAGAGACGATCTGCACCGCAAGTCTTAAAGAAACATTTGTAGAAAC 1020
Db 344 GlyAsnGlyIleGluGluThrIleCysThrAlaSerAlaIleGlyAsnIleGlyArgAsn 363
Qy 1021 AGATCCCTGCATCTACAACCCCTCAAGCTGGTTTCACTCAAACTGCCAACGATCTCAAC 1080
Db 364 ArgSerProAspIleTyrAsnProGlnAlaGlySerLeuIleThrAlaAsnAspLeuAsn 383
Qy 1081 CTTCTAATACTAGTGGCTTGGACCTAGTCTGAATATGAAATCTCTACAGGAATGCA 1140
Db 384 LeuLeuIleLeuArgTyrLeuGlyProSerAlaGluTyrGlyAsnLeuTyrArgAsnAla 403
Qy 1141 TTGTTTTCGCTCACTACAACCAACGACGACAGCATCATATATCATGATTCAGTGGGACGG 1200
Db 404 LeuPheValAlaHisTyrAsnThrAsnAlaHisSerIleIleTyrArgLeuArgGlyArg 423
Qy 1201 GCTCAGCTGAAGTCGTGGACAGCAACGCGCAAGAGTGTACGACGAGGAGCTTCAAGAG 1260
Db 424 AlaHisValGlnValAlaAspSerAsnGlyAsnArgValTyrAspGluGluLeuGlnGlu 443
Qy 1261 GGTCACTGCTTGTGTGTCACAGAACTTCGCGCTCGCTGGAAAGTCCCGAGCGAGAAC 1320
Db 444 GlyHisValLeuValProGlnAsnPheAlaValAlaGlyLysSerGlnSerGluAsn 463
Qy 1321 TTGGAATACGTGGCATTCAGACAGACTCAAGGCCAGCATAGCCAACTCCGCGGTGAA 1380
Db 464 PheGluTyrValAlaPheLysThrAspSerArgProSerIleAlaAsnLeuAlaGlyGlu 483
Qy 1381 AACTCCGCTCATAGATAACCTGCGGAGGAGTGTTCGAAATTCATATGCGCTCCCAAGG 1440
Db 484 AsnSerValIleAspAsnLeuProGluGluValValAlaAsnSerTyrGlyLeuGlnArg 503
Qy 1441 GAGCAGGCAAGCGCAGCTTAAGAAACAACACCCCTTCAAGTTCTTCGTTCCACCGTCTCAG 1500
Db 504 GluGlnAlaArgGlnLeuLysAsnAsnAsnProPheLysPhePheValProProSerGln 523
Qy 1501 CAGTCTCCGAGGCTGTGGCT 1521
Db 524 GlnSerProArgAlaValAla 530
```

RESULT 7

```
US-10-899-551-58
; Sequence 58, Application US/10899551
; Publication No. US20050063994A1
; GENERAL INFORMATION:
; APPLICANT: Caplan, Michael J.
; APPLICANT: Burks, A. Wesley
; APPLICANT: Sampson, Hugh A.
; APPLICANT: Howard, Sossin B.
; APPLICANT: Bottomly, Kim H.
; TITLE OF INVENTION: Methods and Reagents for Decreasing Clinical Reaction to Allergy
; FILE REFERENCE: 2002834-0233
; CURRENT APPLICATION NUMBER: US/10/899,551
; CURRENT FILING DATE: 2004-07-26
```

```
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 58
; LENGTH: 351
; TYPE: PRT
; ORGANISM: species Arachis hypogea
US-10-899-551-58
```

```
Alignment Scores:
Pred. No.: 1,3e-134 Length: 351
Score: 1665.00 Matches: 315
Percent Similarity: 97.83% Conservative: 0
Best Local Similarity: 97.83% Mismatches: 7
Query Match: 61.87% Indels: 0
DB: 17 Gaps: 0
```

US-10-728-051-3 (1-1524) x US-10-899-551-58 (1-351)

```
Qy 1 CGGCAGCAACCGGAGGAGACGCGTCCAGTTCACGGCTCAATCGCAGACCTGAC 60
Db 15 ArgGlnGlnProGluGluAsnAlaCysGlnPheGlnArgLeuAsnAlaGlnArgProAsp 34
Qy 61 AATCGCATTTGAATCAGAGGCGGTTACATTTGAGACTTTGGAAACCCCAACCAACAGGAGTTC 120
Db 35 AsnArgIleGluSerGluGlyGlyTyrIleGluThrTrpAsnProAsnAsnGlnGluPhe 54
Qy 121 GAATCGCGCGGCTCGCCCTCTCTCGTTAGTCTCTCCGCGCAACGCCCTTCGTAGGCCT 180
Db 55 GluCysAlaGlyValAlaLeuSerArgLeuValLeuArgArgAsnAlaLeuArgArgPro 74
Qy 181 TTCTACTCCAACTCTCCCGAGGAGATCTTCATCCAGCAAGGAAGGGGATACTTTGGGTTG 240
Db 75 PheTyrSerAsnAlaProGlnGluIlePheIleGlnGlnGlyArgGlyTyrPheGlyLeu 94
Qy 241 ATATTCCCTGGTTGCTCTAGACACTATGAAGAGCCCTCACACAAAGTCTCGATCTCAG 300
Db 95 IlePheProGlyCysProSerThrTyrGluGluProAlaGlnGlnGlyArgArgSerGln 114
Qy 301 TCCAAAGACCAACCAAGACGTCTCCAAAGGAGAAAGACCAAGCCAAACAGCAACGAGATAGT 360
Db 115 SerGlnArgProProIleArgLeuGlnGlyGluAspGlnSerGlnGlnArgAspSer 134
Qy 361 CACCAAGAGTGCACCGTTTCGATGAGGTGATCTCATTCAGTTCACCGCGGTGTGCT 420
Db 135 HisGlnLysValHisArgPheAspGluGlyAspLeuIleAlaValProThrGlyValAla 154
Qy 421 TTCTGCTCTACAAACCAACCAACGACACTGATGTTGTTGTTGTTCTTCTTACTGACCAAC 480
Db 155 PheTrpLeuTyrAsnAspHisAspThrAspValValAlaValSerLeuThrAspThrAsn 174
Qy 481 AACACGCAACACCAAGCTTGATCAGTTCCCGAGGAGATTCAATTTGGCTGGGAAACAGGAG 540
Db 175 AsnAsnAspAsnGlnLeuAspGlnPheProArgArgPheAsnLeuAlaGlyAsnHisGlu 194
Qy 541 CAAGAGTCTTAAGGTACCGCAACAAAGCAGACAAAGCAGACGAGAGAGAGCTTACCATAT 600
Db 195 GlnGluPheLeuArgTyrGlnGlnGlnSerArgGlnSerArgArgSerLeuProTyr 214
Qy 601 AGCCCATACAGCCCGCAAAAGTACGCTAGACAAAGAGAGCGTCAATTTAGCCCTCGAGGA 660
Db 215 SerProTyrSerProGlnSerGlnProArgGlnGluGluArgPheSerProArgGly 234
Qy 661 CAGCAGACGCGCAGAGAAACGAGCAGACAGAAAGAAAGAAAGAAAGAGTGGAAACATCTTC 720
Db 235 GlnHisSerArgArgGluArgAlaGlyGlnGluGluAsnGluGlyGlyAsnIlePhe 254
Qy 721 AGCGGCTTACGCGCGAGTTCTTGGAAACAAGCCCTTCAGGTTTCAGCAGACAGATAGTG 780
Db 255 SerGlyPheThrProGluPheLeuGluGlnAlaPheGlnValAspAspArgGlnIleVal 274
Qy 781 CAAAACCTAAGAGCGGACCGCAGAGTGAAGAGAGGAGGCCATTGTGACAGTGAGGGGA 840
Db 275 GlnAsnLeuArgGlyGluAsnGluSerGluGluGluGlyAlaIleValThrValArgGly 294
```

QY 841 GGCTCAGAAATCTTGAGCCAGATAGAAAGAGAGCGTCCGACGAAAGAGAGGAATACGAT 900
Db |||||
295 GlyLeuArgIleLeuSerProAspAArgAArgAAspGluGluGluTyrAsp 314
QY 901 GAAGATGAATATGAATACGATGAAGAGGATAGAGCGCTGGCAGGGAAGCAGAGCGAG 960
Db |||||
315 GluAspGluTyrGluTyrAspGluGluAspArgArgGlyArgGlySerArgGlyArg 334
QY 961 GCGAAT 966
Db |||||
335 GlyAsn 336

RESULT 8
US-09-759-967-21
; Sequence 21, Application US/09759967
; Publication No. US20030166518A1
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Nebraska
; TITLE OF INVENTION: METHOD FOR ALLERGEN CHARACTERIZATION
; FILE REFERENCE: UNL 3001.01
; CURRENT APPLICATION NUMBER: US/09/759,967
; CURRENT FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 481
; TYPE: PRT
; ORGANISM: Glycine max
US-09-759-967-21

Alignment Scores:
Pred. No.: 1.69e-120 Length: 481
Score: 1502.00 Matches: 297
Percent Similarity: 71.57% Conservative: 68
Best Local Similarity: 58.24% Mismatches: 91
Query Match: 55.82% Indels: 54
DB: 10 Gaps: 7

US-10-728-051-3 (1-1524) x US-09-759-967-21 (1-481)

QY 1 CGCAGCAACCGGAGGAGACGGTCCAGTTCAGGGCTCAATGGCAGAGACCTGAC 60
Db |||||
23 ArgGluGlnProGlnGlnAsnGluCysGlnIleGlnArgLeuAsnAlaLeuLysProAsp 42
QY 61 AATCGCATTCGAATCAGAGGCGGTTACATTGACACTTGGAAACCCCAACACAGGAGTTC 120
Db |||||
43 AsnArgIleGluSerGluGlyGlyPheIleGluThrTrpAsnProAsnLysProPhe 62
QY 121 GAATGCCCGCGTCCGCTCTCGCTTAGTCTCGCGCAACCGCCCTTCGTAGGCCT 180
Db |||||
63 GlnCysAlaGlyValAlaLeuSerArgCysThrLeuAsnArgAsnAlaLeuArgPro 82
QY 181 TTCTACTCCATGCTCCCGAGGATCTTCATCCAGCAAGGAGGATCTTTGGTTC 240
Db |||||
83 SerTyrThrAsnAlaProGlnGluIleTyrIleGlnGlnGlySerGlyIlePheGlyMet 102
QY 241 ATATTCCCTGGTGTCTCTAGACACTATGAAGAGCTCACACAAAGTCTCGATCTCAG 300
Db |||||
103 IlePheProGlyCysProSerThrPheGluGluProGlnGlnGlyGly-----Gln 119
QY 301 TCCCAAGACACCAACAGACCTCTCCAGGAGAGACCAAGCCAAAGCCAAACAGAGATAGT 360
Db |||||
120 SerSerArgPro-----GlnAspArg 126
QY 361 CACAGAGGTGCACCGCTTCGATGAGGTGATCTCATTCAGTCCACCGGTGTGCT 420
Db |||||
127 HisGlnIleTyrHisPheArgGluGlyAspLeuIleAlaValProThrGlyPheAla 146
QY 421 TTCTGGCTCTACACGACACGACACTGATGTTGTCTGCTCTTCTTACTGACACCAAC 480
Db |||||
147 TyrTrpMetTyrAsnAsnGluAspThrProValValAlaValSerLeuIleAspThrAsn 166

RESULT 9

QY 481 AACCAACGACAAACCAAGCTTGATCAGTCCCGAGGAGATTCAATTTGGCTGGGAACACGAG 540
Db |||||
167 SerPheGlnAsnGlnLeuAspGlnMetProArgArgPheTyrLeuAlaGlyAsnGlnGlu 186
QY 541 CAAGAGTTCTTAAGGTACCAACCAAGCAGACAAACAGCAGACGAAAGAGCTTACCATAT 600
Db |||||
187 GlnGluPheLeuGlnTyrGlnProGlnLysGlnGlnGly----- 199
QY 601 AGCCCATACAGCCCGCAAAAGTCAGCTAGACAAAGAGCGTGAATTTAGCCCTCGAGGA 660
Db |||||
200 -----Gly 200
QY 661 CAGCACACCCGACAGAGAACGACGAGGACAAAGAAACGAGGTGGAAAACATCTTC 720
Db |||||
201 ThrGlnSerGlnLysGlyArgGlnGlnGluGluAsnGluGlyGlySerIleLeu 220
QY 721 AGCGGCTTCACGCGGAGTCTCTGGAACAAGCCTTCCAGGTTGACGACGACAGACATG 780
Db |||||
221 SerGlyPheAlaProGluPheLeuGluHisAlaPheValVal---AspArgGlnIleVal 239
QY 781 CAAAACCTTAAGAGCGGACGACGAGAGTGAAGAAGAGGAGGAGCCATTGTGACAGTGAGGGA 840
Db |||||
240 ArgLysLeuGlnGlyGluAsnGluGluGluGlyGlyAlaIleValThrValLysGly 259
QY 841 GGCCTCAGAAATCTTGAGCCCA-----GATGAAGAAGACAGCTGCCGACGAGAAAG 891
Db |||||
260 GlyLeuSerValIleSerProThrGluGlnGlnGlnArgProGluGluGlu 279
QY 892 GAATACGATGAAGATGAATATCAATACGATGAAGAGGATAGAGCGCTGGCAGGGAAGC 951
Db |||||
280 LysProAsp-----CysAspGluLysAspLysHisCysGlnSerGlnSer 294
QY 952 AGAGCAGCGGGGAATGGTATTGAAGAGACGATCTCCACCGCAAGTCTTAAAGAACATT 1011
Db |||||
295 Arg-----AsnGlyIleAspGluThrIleCysThrMetArgLeuArgHisAsnIle 311
QY 1012 GGTAGAAACAGATCCCTGACATCTACAACCTCAGCTGGTTCATCTCAAACTGCCAAC 1071
Db |||||
312 GlyGlnThrSerSerProAspIlePheAsnProGlnAlaGlySerIleThrAlaThr 331
QY 1072 GATCTCAACTTCTAATATCTAGTGGCTTGACCTAGTGTGTAATATAGAAATCTTAC 1131
Db |||||
332 SerLeuAspPheProAlaLeuSerTrpLeuLysLeuSerAlaGlnPheGlySerLeuArg 351
QY 1132 AGGAATGCAATGTTTGTGCTCCTACCAACCAACGACACGACATCATATATCGATTG 1191
Db |||||
352 LysAsnAlaMetPheValProHisTyrAsnLeuAsnAlaAsnSerIleIleTyrAlaLeu 371
QY 1192 AGGGACGGCTCAGTCGACAGTCTGGACCAAGCGCAACAGAGTGTACACGAGGAG 1251
Db |||||
372 AsnGlyArgAlaLeuValGlnValValAsnCysAsnGlyGluArgValPheAspGlyGlu 391
QY 1252 CTTCAAGAGGGTCAGCTGCTTGTGTGCGCACAGAACTTCGCGCTCGCTGGAAAGTCCAG 1311
Db |||||
392 LeuGlnGluGlyGlnValIleValProGlnAsnPheAlaValAlaAlaArgSerGln 411
QY 1312 AGCGAACTTCGAATACGTGCGCATTCAGACAGACTCAAGCCCGACGATACCCACCTC 1371
Db |||||
412 SerAspAsnPheGluTyrValSerPheLysThrAsnAspArgProSerIleGlyAsnLeu 431
QY 1372 GCGGTGAAACTCCGCTCATATCACTACCTGCGGAGGAGTGTGTCAAATTCATATGCG 1431
Db |||||
432 AlaGlyAlaAsnSerLeuLeuAsnAlaLeuProGluGluValIleGlnGlnThrPheAsn 451
QY 1432 CTCCAAGGAGGACGACGAGGAGCTTAAAGAACAAACCCCTTCAAGTCTTCGTTCCA 1491
Db |||||
452 LeuArgArgGlnGlnAlaArgGlnValLysAsnAsnProPheSerPheLeuValPro 471
QY 1492 CGGTCTCAGCAGTCTCCGAGGCGCTGSGCT 1521
Db |||||
472 ProLysGluSerGlnArgArgValValAla 481

RESULT 10
US-10-425-114-43865
; Sequence 43865, Application US/10425114
; Publication No. US20040034889A1
; GENERAL INFORMATION:
; APPLICANT: Iiu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovacic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E

APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(5313)B
CURRENT APPLICATION NUMBER: US/10/425,114
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 43865
LENGTH: 488
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: 700645818_FLI.pep
US-10-425-114-43865

Alignment Scores:
Pred. No.: 1 7e-120 Length: 488
Score: 1502.00 Matches: 297
Percent Similarity: 71.57% Conservative: 68
Best Local Similarity: 58.24% Mismatches: 91
Query Match: 55.82% Indels: 54
DB: 15 Gaps: 7

US-10-728-051-3 (1-1524) x US-10-425-114-43865 (1-488)

QY 1 CGGCAGCAACCGGAGGAGACGGCTCCAGTTCAGCGCTCAATGCGCAGAGACCTGAC 60
DB 30 ArgGluGlnProGlnGlnAsnGluCysGlnIleGlnArgLeuAsnAlaLeuLysProAsp 49
QY 61 AATCGATTGAATCAGAGGCGGTTCATTAGACTTGAACCCCAACACGAGGATTC 120
DB 50 AsnArgIleGluSerGluGlyGlyPheIleGluThrTrpAsnProAsnAsnLysProPhe 69
QY 121 GAATCGCGCGCTCCCTCTCTCTAGTCTCGCGCGCAACGCCCTTCGTAGCCT 180
DB 70 GlnCysAlaGlyValAlaLeuSerArgCysThrLeuAsnArgAsnAlaLeuArgPro 89
QY 181 TTCTACTCAATCTCCCGAGGAGATCTTCATCCAGCAAGGAGGATCTTTGGGTG 240
DB 90 SerTyrThrAsnAlaProGlnGluIleTyrIleGlnGlnGlySerGlyIlePheGlyMet 109
QY 241 ATATTCCTGGTGTCTGACACTATGAGAGCTTCACACAAAGCTCGTCGATCTCAG 300
DB 110 IlePheProGlyCysProSerThrPheGluGluProGlnGlnLysGly-----Gln 126
QY 301 TCCCAAGACCAACAGAGCTCTCCAGGAGAGACCAAGCAACGAGATAGT 360
DB 127 SerSerArgPro-----GlnAspArg 133
QY 361 CACCAAGAGTGACCGGTTTCGATGAGGTGATCTCATTTGCAGTTCACCGCGGTGTGCT 420
DB 134 HisGlnLysIleTyrHisPheArgGluGlyAspLeuIleAlaValProThrGlyPheAla 153
QY 421 TTCTGGCTTACACGACCAACGACACTGATGTTGTTGCTTCTTACTGACACCAAC 480
DB 154 TyrTrpMetTyrAsnAsnGluAspThrProValValAlaValSerLeuIleAspThrAsn 173
QY 481 AACCAACGACCAACGCTGATCAGTTCCTCCAGGAGATTCATTTGCTGGGACACGGAG 540
DB 174 SerPheGlnAsnGlnLeuAspGlnMetProArgArgPheTyrLeuAlaGlyAsnGlnGlu 193
QY 541 CAAGAGTCTTAAAGTACCAAGCAACAAAGCAGACAAAGCAGACGAAAGCTTACCATAT 600
DB 194 GlnGluPheLeuGlnTyrGlnProGlnLysGlnGlnGly----- 206
QY 601 AGCCCATACAGCCCGCAAGTACGCTCAGCTAGACAAAGAGCGTGAATTTAGCCCTCAGGA 660
DB 207 -----Gly 207
QY 661 CAGCAGCGCAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 720
DB 208 ThrGlnSerGlnLysGlyLysArgGlnGlnGlnGluAsnGlnGlySerIleLeu 227

QY 721 AGCGGCTTCACGCGGAGTTCCTGGAACAAGCCTTCAGGTTGACGACAGACAGATAGTG 780
DB 228 SerGlyPheAlaProGluPheLeuGluHisAlaPheValVal---AspArgGlnIleVal 246
QY 781 CAAAACCTTAAGAGGCGGAGACCGAGAGTGAAGAAGAGGAGCCATTGTGACAGTGGGA 840
DB 247 ArgLysLeuGlnGlyGluAsnGluGluGluGlyLysGlyAlaIleValThrValLysGly 266
QY 841 GGCCTCAGAACTTGTAGCCCA-----GATAGAAGAGAGCGTCCCGACCAAGAGAG 891
DB 267 GlyLeuSerValIleSerProThrGluGlnGlnGlnArgProGluGluGluGlu 286
QY 892 GAATACGATCAAGATCAATATCAATACGATGAAGAGGATGAAGGCGTCCGACGGGAAGC 951
DB 287 LysProAsp-----CysAspGluLysAspLysHisCysGlnSerGlnSer 301
QY 952 AGAGCAGGGGAATGGTATTGAAGAGAGTCTGCACCGCAAGTGTCTAAAGAACATT 1011
DB 302 Arg-----AsnGlyIleAspGluThrIleCysThrMetArgLeuArgHisAsnIle 318
QY 1012 GGTAGAAACAGATCCCTGACATCTACACCTCAGCTGCTCACTCAAAACTGCCAAC 1071
DB 319 GlyGlnThrSerSerProAspIlePheAsnProGlnAlaGlySerIleThrThrAlaThr 338
QY 1072 GATCTCAACCTTCTAATACTTAGTGGCTTGACCTAGTGTCTCAATATGAAATCTCTAC 1131
DB 339 SerLeuAspPheProAlaLeuSerTrpLeuLysLeuSerAlaGlnPheGlySerLeuArg 358
QY 1132 AGAATGCAATTTGTTGCTCCTCACTAACACCAACGACGACGACGACGACGACGACGACG 1191
DB 359 LysAsnAlaMetPheValProHisTyrAsnLeuAsnAlaAsnSerIleIleTyrAlaLeu 378
QY 1192 AGGGACGCGCTCAGTGCAGCTGCTGGACACCAACGACGAGTGTACGACGAGGAG 1251
DB 379 AsnGlyArgAlaLeuValGlnValValAsnCysAsnGlyGluArgValPheAspGlyGlu 398
QY 1252 CTTCAAGAGGGTCAAGTGTGCTGCTGCGCACAGAACTTCGCGCTGCTGGAAGTCCACG 1311
DB 399 LeuGlnGluGlyGlnValLeuIleValProGlnAsnPheAlaValAlaAlaArgSerGln 418
QY 1312 ACAGAGACTTCGATATGCTGCAATTCAGACAGACTCAAGCCCGACGACGACGACGACGAC 1371
DB 419 SerAspAsnPheGluTyrValSerPheLysThrAsnAspArgProSerIleGlyAsnLeu 438
QY 1372 GCGGTGAAAACCTCGCTCATATACCTCGCGGAGGAGTGTGTCAAATTCATATGSC 1431
DB 439 AlaGlyAlaAsnSerLeuLeuAsnAlaLeuProGluGluValIleGlnGlnThrPheAsn 458
QY 1432 CTCCAAAGGAGCAGGACGAGCAGCTTAAGAAACAACACCCCTTCAAGTTCCTGTTCCA 1491
DB 459 LeuArgArgGlnGlnAlaArgGlnValLysAsnAsnProPheSerPheLeuValPro 478
QY 1492 CCGTCTCAGCAGTCTCCGAGGCGTGTGCT 1521
DB 479 ProLysGluSerGlnArgValValAla 488

RESULT 11

US-10-425-114-44047
; Sequence 44047, Application US/10425114
; Publication NO. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Zhou, Yihong
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128

Alignment Scores:

Pred. No.: 1.7e-120 Length: 492
 Score: 1502.00 Matches: 297
 Percent Similarity: 71.57% Conservative: 68
 Best Local Similarity: 58.24% Mismatches: 91
 Query Match: 55.82% Indels: 54
 DB: 15 Gaps: 7

US-10-728-051-3 (1-1524) x US-10-425-114-43836 (1-492)

```

QY 1 CGGCACACCGGAGGAGACGGCTCCAGGCTCAATCGGCAGAGACTGAC 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 34 ArgGluGlnProGlnGlnAsnGluCysGlnIleGlnArgLeuAsnAlaLeuLysProAsp 53
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 AATCGCATTTGAATCAGAGGCGGTTTACATTGAGACTTGGAAACCCCAACAGGAGTTC 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 54 AsnArgIleGluSerGluGlyPheIleGluThrTrpAsnProAsnAsnLysProPhe 73
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 GAATCGCGGCTCGCCCTCTCTCGCTTAGTCTCCCGCGCAACGCCCTTCGTAGGCT 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 74 GlnCysAlaGlyValAlaLeuSerArgCysThrLeuAsnArgAsnAlaLeuArgPro 93
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 181 TTCTACTCCATGCTCCCGAGGATCTTCATCCAGCAAGGAGGAGTACTTTGGGTTG 240
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 94 SerTyrThrAsnAlaProGlnGluIleTyrIleGlnGlnGlySerGlyIlePheGlyMet 113
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 241 ATATTCCCTGTTGCTCAGACACTATGAAGAGCCTCACACACAGGTCTCGATCTCAG 300
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 114 IlePheProGlyCysProSerThrPheGluGluProGlnGlnLysGly-----Gln 130
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 301 TCCCAAGACCCCAAGAGCTCTCCAGGAGAGACCAAGCCAAAGCAGCAACGAGATAGT 360
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 131 SerSerArgPro-----GlnAspArg 137
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 361 CACGAGAGGTGCACCGTTTCGATGAGGTGATCTCATTTGCAGTTCACCGGTTGCT 420
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 138 HisGlnLysIleTyrHisPheArgGluGlyAspLeuIleAlaValProThrGlyPheAla 157
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 421 TTCTGGCTCTACACGACGACGACACTGATGTGTTGTTGTTCTTCTTACTGACACCAAC 480
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 158 TyrTrpMetTyrAsnAsnGluAspThrProValValAlaValSerLeuIleAspThrAsn 177
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 481 AACACGACACACGAGTTGATCTCCCGAGAGATTCATTTGCTGGGACACGGAG 540
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 178 SerPheGlnAsnGlnLeuAspGlnMetProArgArgPheTyrLeuAlaGlyAsnGlnGlu 197
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 541 CAAGAGTTCTTAGGTACCGACCAACAAAGCAGCAACAGCAGCAAGAGAGCTTACCATAT 600
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 198 GlnGluPheLeuGlnTyrGlnProGlnLysGlnGlnGly-----210
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 601 AGCCCATACAGCCCGCAAGTACGCTAGACAAAGAGAGCGGTGAATTTAGCCCTCGAGGA 660
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 211 -----Gly 211
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 661 CAGCAGACGCGCAGAGACGAGCAGCAGCAAGAGAGAAACAAAGGTGGAACATCTTC 720
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 212 ThrGlnSerGlnLysGlyLysArgGlnGlnGluGluAsnGluGlySerIleLeu 231
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 721 AGCGGCTTCAGCGCGGAGTCTCGAACACAGCCCTTCAGGTTACACGACAGATAGTG 780
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 232 SerGlyPheAlaProGluPheLeuGluHisAlaPheValVal---AspArgGlnIleVal 250
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 781 CAAACCTTAAGCGCAGACCGGAGAGTGAAGAGAGGAGGCCATTTGACAGTGGAGGA 840
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 251 ArgLysLeuGlnGlyGluAsnGluGluGluGlyAlaIleValThrValLysGly 270
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 841 GGCTCTCAGATCTTGACCCA-----GATAGAAAGAGAGCTGCCGACGAAGAAGAG 891
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 271 GlyLeuSerValIleSerProThrGluGluGlnGlnArgProGluGluGluGlu 290
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 892 GAATACGATCAAGATGAATATGATACGATGAGAGAGATAGAAGCGGTGGCAGGGAAGC 951
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 291 LysProAsp-----CysAspGluLysAspLysHisCysGlnSerGlnSer 305
QY 952 AGAGCAGCGGGGAATGTTAAGAGACGATCTCCACCGCAAGTGCTAAAGAACATT 1011
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 306 Arg-----AsnGlyIleAspGluThrIleCysThrMetArgLeuArgHisAsnIle 322
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1012 GGTAGAAACAGATCCCTCTGACATCTACAAACCTCAAGCTGGTTCACTCAAACTGCCAAC 1071
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 323 GlyGlnThrSerSerProAspIlePheAsnProGlnAlaGlySerIleThrThrAlaThr 342
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1072 GATCTCAACCTTCTAATCTAGTGGCTTGGACCTAGTCTGAATATGGAATCTCTAC 1131
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 343 SerLeuAspPheProAlaLeuSerTrpLeuLysLeuSerAlaGlnPheIleSerLeuArg 362
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1132 AGGAATGATGATTGTTGTGCTCAGTCAACACCAACGACACAGCATCATATCGATTG 1191
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 363 LysAsnAlaMetPheValProHisTyrAsnLeuAsnAlaAsnSerIleIleTyrAlaLeu 382
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1192 AGGACGCGGCTCAGTGCAGTCTGGACAGCAACGCAACAGAGTGTACACGAGGAG 1251
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 383 AsnGlyArgAlaLeuValGlnValValAsnCysAsnGlyGluArgValPheAspGlyGlu 402
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1252 CTTCAAGAGGCTCAGTGTGCTGTCACAGAACTTCGCCGCTCGTGAAGTCCACAG 1311
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 403 LeuGlnGluGlyGlnValLeuIleValProGlnAsnPheAlaValAlaArgSerGln 422
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1312 AGCGAAGACTTCGAATCTGCGCATTCAGACAGACTCAAGGCCAGCATAGCCCACTC 1371
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 423 SerAspAsnPheGluTyrValSerPheLysThrAsnAspArgProSerIleGlyAsnLeu 442
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1372 GCCGTGAAACTCCGTCATAGATAACCTGCCGAGGAGGTGGTTCGCAATTCATATGGC 1431
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 443 AlaGlyAlaAsnSerLeuLeuAsnAlaLeuProGluGluValIleGlnGlnThrPheAsn 462
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1432 CTCCTCAAGGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1491
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 463 LeuArgArgGlnGlnAlaArgGlnValLysAsnAsnAsnProPheSerPheLeuValPro 482
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1492 CGCTCTCAGCAGTCTCCGAGGCGCTGGCT 1521
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 483 ProLysGluSerGlnArgArgValValAla 492
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 13

US-10-425-114-46294

; Sequence 46294, Application US/10425114

; Publication No. US20040034888A1

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E

; APPLICANT: Tabaska, Jack E

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with

; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 46294

; LENGTH: 492

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: 700757865_FLI.pep

US-10-425-114-46294

Alignment Scores:

Pred. No.: 1.7e-120 Length: 492
 Score: 1502.00 Matches: 297
 Percent Similarity: 71.57% Conservative: 68
 Best Local Similarity: 58.24% Mismatches: 91
 Query Match: 55.82% Indels: 54

```
DB: 15 Gaps: 7
US-10-728-051-3 (1-1524) x US-10-425-114-46294 (1-492)
QY 1 CGGCAGCAACCGGAGGAGACGGTTCAGCTTCAGCGCTCAATGCGCAGACCTGAC 60
Db 34 ArgGluGlnProGlnAenGluCysGlnIleGlnArgLeuAenAlaLeuLysProAsp 53
QY 61 AATCGCATTAATCAGAGCGGCTTACATTGAGACTTGGAAACCCCAACACAGAGATTC 120
Db 54 AsnArgIleGluSerGluGlyGlyPheIleGluThrTrpAsnProAenLysProPhe 73
QY 121 GAATGCGCGGCTCGCCCTCTCTCGCTAGTCTCCGCGCAACGCCCTTCGTAGGCT 180
Db 74 GlnCysAlaGlyValAlaLeuSerArgCysThrLeuAenArgAenAlaLeuArgPro 93
QY 181 TTCTACTCCAAATGCTCCCAAGAGATCTTCATCCAGCAAGGAGGATCTTTGGGTG 240
Db 94 SerTyrThrAenAlaProGlnGluIleTyrIleGlnGlnGlySerGlyIlePheGlyMet 113
QY 241 ATATTCCCTGGTTGCTTAGACACTATGAAGAGCTCACAACAAGTCTGCTGATCTCAG 300
Db 114 IlePheProGlyCysProSerThrPheGluGluProGlnGlnLysGly-----Gln 130
QY 301 TCCCAAGACCAACAGACTCTCCAGGAGAGACCAAGCCAAACAGCAACGAGATAGT 360
Db 131 SerSerArgPro-----GlnAspArg 137
QY 361 CACCAAGAGTGCAACCGTTTCGATAGGCTGATCTCATTGCAGTTCACCGGCTGTGCT 420
Db 138 HisGlnLysIleTyrHisPheArgGluGlyAspLeuIleAlaValProThrGlyPheAla 157
QY 421 TTCTGCTCTTACACGACCAACGACACTGATGTTGTTGCTGTTCTTCTACTGACCAAC 480
Db 158 TyrTrpMetTyrAenAenGluAepThrProValValAlaValSerLeuIleAepThrAen 177
QY 481 AACCAACGACACAGCTTGATGATGTTCCCGAGGAGATTCATTTGCTGGGACACGGAG 540
Db 178 SerPheGlnAenGlnLeuAepGlnMetProArgArgPheTyrLeuAlaGlyAenGlnGlu 197
QY 541 CAAGAGTCTTAAGGTACCAACGACCAACAAAGCAGACGACGACGACGACGACGATAT 600
Db 198 GlnGluPheLeuGlnTyrGlnProGlnLysGlnGlnGly-----Gly 211
QY 601 AGCCCATACAGCCCGCAAGTACGCTTAGACAAGAGAGCGTGAATTTAGCCCTCAGGA 660
Db 211 -----Gly 211
QY 661 CAGCAGCCGACAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 720
Db 212 ThrGlnSerGlnLysGlyTyrArgGlnGlnGlnGluGluAenGluGlyGlySerIleLeu 231
QY 721 AGCGGCTTACCGCGGAGTTCCTGGAAACAGCTTCAGGTTGACGACGACGACGATAGTG 780
Db 232 SerGlyPheAlaProGluPheLeuGluHieAlaPheValVal-----AspArgGlnIleVal 250
QY 781 CAAACCTTAAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 840
Db 251 ArgLysLeuGlnGlyGluAenGluGluGluGluLysGlyAlaIleValThrValLysGly 270
QY 841 GGCTCGAATCTTGACCCCA-----GATAGAAAGACGCTGCCGACGAAAGAGAG 891
Db 271 GlyLeuSerValIleSerProThrGluGlnGlnGlnArgProGluGluGluGlu 290
QY 892 GAATACGATGAAGTGAATATGATGATGAAGAGATAGAGGCGTGGCAGGGGAGC 951
Db 291 LysProAsp-----CysAspGluLysAspLysHisCysGlnSerGlnSer 305
QY 952 AGAGGACGGGGAGTGTATTGAGAGACGATCTGCACCGCAAGTGTCTAAAGAAACATT 1011
Db 306 Arg-----AenGlyIleAepGluThrIleCysThrMetArgLeuArgHisAenIle 322
QY 1012 GGTAGAAACAGATCCCTGATCTACACCCCTCAAGCTGGTTCACTCAAAAACCTGCAAC 1071
```

```
Db 323 GlyGlnThrSerSerProAspIlePheAenProGlnAlaGlySerIleThrAlaThr 342
QY 1072 GATCTCAACTTCTTAATACTTAGGTGGCTTAGGACCTAGTGGTAATATGAAATCTCTAC 1131
Db 343 SerLeuAepPheProAlaLeuSerTrpLeuLysLeuSerAlaGlnPheGlySerLeuArg 362
QY 1132 AGAATGCAATTTGTTGCTCCTCATTACACACCAACGACGACGACGATCATATATCGATTG 1191
Db 363 LysAenAlaMetPheValProHisTyrAenLeuAenAlaAenSerIleIleTyrAlaLeu 382
QY 1192 AGGGGACGGCTCACGTGCAAGTCGTGGACAGCAACGGCAACAGAGTGTACGACGAGGAG 1251
Db 383 AsnGlyArgAlaLeuValGlnValValAenCysAenGlyGluArgValPheAspGlyGlu 402
QY 1252 CTTCAAGAGGGTCACTGCTTGTGGTCCACAGAACTTCGCGCTCGCTGGAAAGTCCCGAG 1311
Db 403 LeuGlnGluGlyGlnValLeuIleValProGlnAenPheAlaValAlaAArgSerGln 422
QY 1312 AGGAGAACTTTCGAATACGTGGCATTCAGACAGACTCAAGGCCGACGATAGCCAACTC 1371
Db 423 SerAspAenPheGluTyrValSerPheLysThrAenAspArgProSerIleGlyAenLeu 442
QY 1372 GCGGTGAAACTCCGTCTATAGTAACTCCGCGGAGGAGTGTTCCTCAAAATTCATATGC 1431
Db 443 AlaGlyAlaAenSerLeuAenAlaLeuProGluGluValIleGlnGlnThrPheAen 462
QY 1432 CTCCAAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1491
Db 463 LeuArgArgGlnGlnAlaArgGlnValLysAenAenAenProPheSerPheLeuValPro 482
QY 1492 CGGTCTCAGCAGTCTCCGAGGCGTGTGGCT 1521
Db 483 ProLysGluSerGlnArgArgValValAla 492
```

RESULT 14

```
US-10-425-114-71959
; Sequence 71959, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 71959
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700676992_FLI pep
US-10-425-114-71959
```

```
Alignment Scores:
Pred. No.: 2,74e-120 Length: 466
Score: 1499.50 Matches: 297
Percent Similarity: 72.02% Conservative: 71
Best Local Similarity: 58.12% Mismatches: 94
Query Match: 55.72% Indels: 49
DB: 15 Gaps: 6
```

US-10-728-051-3 (1-1524) x US-10-425-114-71959 (1-466)

```
QY 1 CGGCAGCAACCGGAGGAGACGGTTCAGCTTCAGCGCTCAATGCGCAGACCTGAC 60
Db 1 ArgGluGlnAlaGlnGlnAenGluCysGlnIleGlnLysLeuAenAlaLeuLysProAsp 20
```

```
QY 61 AATCGCATTAATCAGAGCGCGTTACATTGAGACTTGAACCCCAACACAGGAGTTC 120
Db 21 AsnArgIleGluSerGluGlyGlyPheIleGluThrTrpAsnProAsnAsnLysProPhe 40
QY 121 GAATGCGCGGGTCCGCTCTCTCGCTTAGCTTCCTCCGCGGCAAGCCCTTCGTAGGCT 180
Db 41 GlnCysAlaGlyValAlaLeuSerArgCysThrLeuAsnArgAsnAlaLeuArgPro 60
QY 181 TTCTACTCAATGCTCCCGAGAGATCTTCATCCAGCAAGGAGGGGACTTTGGGTG 240
Db 61 SerTyrAsnGlyProGlnGluIleTyrIleGlnGlnGlyAsnGlyIlePheGlyMet 80
QY 241 ATATTCCTCGGTTCCTAGACACTATGAAGAGCCTCACACAAAGGTGCTGATCTCAG 300
Db 81 IlePheProGlyCysProSerThrTyrGlnGluProGlnGluSerGlnGlnArgGlyArg 100
QY 301 TCCCAAGACCAACCAAGAGCTCTCCAGAGGAGAACCAAGCCACAGCAAGAGATAGT 360
Db 101 SerGlnArgPro-----GlnAspArg 107
QY 361 CACCAAGAGGTGCACCGTTTCGATGAGGTGATCTCATTTGAGTTCCTCCACCGGTGTGCT 420
Db 108 HisGlnLysValHisArgPheArgGluGlyAspLeuIleAlaValProThrGlyValAla 127
QY 421 TTCTGGCTCTACCAAGCACCAAGCACTGATGTTGTTGCTGTTTCTCTTACTGACACCAAC 480
Db 128 TrpTyrMetTyrAsnAsnGluAspThrProValValAlaValSerIleIleAspThrAsn 147
QY 481 AACCAAGCAACAGCTGATGATGTTCCCGAGAGATTCATTTGCTGGGCAACAGCGAG 540
Db 148 SerLeuGluAsnGlnLeuAspGlnMetProArgArgPheTyrLeuAlaGlyAsnGlnGlu 167
QY 541 CAAGAGTCTTAAGTACCAAGCAACAAAGCAGACAAAGCAGACGAAGAAGCTTACCATAT 600
Db 168 GlnGluPheLeuLysTyrGlnGlnGln-----GlnAspArg 176
QY 601 AGCCCATACAGCCGCAAGTACGCTAGCAAGAGAGAGCGTGAATTTAGCCCTCGAGGA 660
Db 177 -----GlnGlnGlySerGlnSerGlnLysGly 186
QY 661 CAGCAGCCGCGAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 720
Db 187 LysGln-----GlnGluGluAsnGluGlySerAsnIleLeu 199
QY 721 AGCGGCTTACGCGGAGTTCCTGGAACAGCGCTTCAGGTTGACGACAGCAGATAGTG 780
Db 200 SerGlyPheAlaProGluPheLeuLysGluAlaPheGlyVal---AsnMetGlnIleVal 218
QY 781 CAAACCTTAAGCGCGAGACCGAGAGTGAAGAGAGGAGCCATTTGACAGTGAAGGGA 840
Db 219 ArgAsnLeuGlnGlyGluAsnGluGluGluAspSerGlyAlaIleValThrValLysGly 238
QY 841 GGCTCAGATCTTGACCCGAGATAGAAAGAGAGCGTCCCGACAGAGAGGAGATACGAT 900
Db 239 GlyLeuArgValThrAlaProAlaMetArgLysProGlnGlnGluAspAspAsp 258
QY 901 GAAGATGAA-----TATGAATACCATGAAGAGGATAGAGCGCTGGCAGCGGGA 948
Db 259 GluGluGlnProGlnCysValGluThrAspLysGlyCysGlnArgGlnSerLysArg 278
QY 949 AGCAGAGCGAGGGGAATGTTGTAAGAGACGATCTGACCCGCAAGTGTCTAAAGAAC 1008
Db 279 SerArg-----AsnGlyIleAspGluThrIleCysThrMetArgLeuArgGlnAsn 295
QY 1009 ATTGGTAGAACACATCCCTGATCTACAACTCAAGCTGTTGCTTCACTCAAACTGCC 1068
Db 296 IleGlyGlnAsnSerProAspIleTyrAsnProGlnAlaGlySerIleThrAla 315
QY 1069 AACCATCTCAACCTTCTAATACTTAGTGGCTTGGACCTAGTGTGAATATGGAATCTC 1128
Db 316 ThrSerLeuAspPheProAlaLeuTrpLeuLysLeuSerAlaGlnTyrGlySerLeu 335
```

```
QY 1129 TACAGGAATGATTGTTTGTGCTCACTACAAACCAACGACGACGATCATATATCGA 1188
Db 336 ArgLysAsnAlaMetPheValProHisTyrThrLeuAsnAlaAsnSerIleIleTyrAla 355
QY 1189 TTGAGGGACGGGCTCAGTGTCAAGTGTGACAGCAACGCGCAACAGAGTGTACGACGAG 1248
Db 356 LeuAsnGlyArgAlaLeuValGlnValValAsnCysAsnGlyGluArgValPheAspGly 375
QY 1249 GAGCTTCAAGAGGCTCAGTGTCTTGTGTGCTCCACAGAACTTCGCGCTGCGTGAAGTCC 1308
Db 376 GluLeuGlnGluGlyGlyValLeuIleValProGlnAsnPheAlaValAlaAlaLysSer 395
QY 1309 CAGACGAGAACTTCGATAGTGTGCTTCAAGCAGACTCAAGCGCCAGCAGATCCCAAC 1368
Db 396 GlnSerAspAsnPheGluTyrValSerPheLysThrAsnAspArgProSerIleGlyAsn 415
QY 1369 CTCGCGGTGAAAACCTCCGTATAGATAAACCCTGCGGAGGAGGTGGTTCGAAATTCATAT 1428
Db 416 LeuAlaGlyAlaAsnSerLeuLeuAsnAlaLeuProGluGluValIleGlnHisThrPhe 435
QY 1429 GGCCTCCAAAGGAGCAGCGCAAGCAGCTTAAAGAAACAACAACCCCTTCAAGTTCCTGTT 1488
Db 436 AsnLeuLysSerGlnGlnAlaArgGlnValLysAsnAsnAsnProPheSerPheLeuVal 455
QY 1489 CCACGCTCTCAGCAGCTCTCCGAGGGCTGTGGCT 1521
Db 456 ProProGlnGluSerGlnArgArgAlaValAla 466

RESULT 15
US-10-425-114-71973
; Sequence 71973, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 71973
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700677092_FLI.pep
US-10-425-114-71973

Alignment Scores:
Pred. No.: 2,74e-120 Length: 466
Score: 1499.50 Matches: 297
Percent Similarity: 72.02% Conservatives: 71
Best Local Similarity: 58.12% Mismatches: 94
Query Match: 55.72% Indels: 49
DB: 15 Gaps: 6

US-10-728-051-3 (1-1524) x US-10-425-114-71973 (1-466)
QY 1 CGGCAGCAACCGAGAGAGAACCGCTGCCAGTTCACGCGCTCAATGCGCAGAGACCTGAC 60
Db 1 ArgGluGlnAlaGlnGlnAsnGluCysGlnIleGlnLysLeuAsnAlaLeuLysProAsp 20
QY 61 AATCGCATTAATCAGAGCGCGTTACATTGAGACTTGAACCCCAACACAGGAGTTC 120
Db 21 AsnArgIleGluSerGluGlyGlyPheIleGluThrTrpAsnProAsnAsnLysProPhe 40
QY 121 GAATGCGCGGGTCCGCTCTCTCGCTTAGCTTCCTCCGCGGCAAGCCCTTCGTAGGCT 180
Db 41 GlnCysAlaGlyValAlaLeuSerArgCysThrLeuAsnArgAsnAlaLeuArgPro 60
QY 181 TTCTACTCAATGCTCCCGAGAGATCTTCATCCAGCAAGGAGGGGACTTTGGGTG 240
Db 61 SerTyrAsnGlyProGlnGluIleTyrIleGlnGlnGlyAsnGlyIlePheGlyMet 80
QY 241 ATATTCCTCGGTTCCTAGACACTATGAAGAGCCTCACACAAAGGTGCTGATCTCAG 300
Db 81 IlePheProGlyCysProSerThrTyrGlnGluProGlnGluSerGlnGlnArgGlyArg 100
QY 301 TCCCAAGACCAACCAAGAGCTCTCCAGAGGAGAACCAAGCCACAGCAAGAGATAGT 360
Db 101 SerGlnArgPro-----GlnAspArg 107
QY 361 CACCAAGAGGTGCACCGTTTCGATGAGGTGATCTCATTTGAGTTCCTCCACCGGTGTGCT 420
Db 108 HisGlnLysValHisArgPheArgGluGlyAspLeuIleAlaValProThrGlyValAla 127
QY 421 TTCTGGCTCTACCAAGCACCAAGCACTGATGTTGTTGCTGTTTCTCTTACTGACACCAAC 480
Db 128 TrpTyrMetTyrAsnAsnGluAspThrProValValAlaValSerIleIleAspThrAsn 147
QY 481 AACCAAGCAACAGCTGATGATGTTCCCGAGAGATTCATTTGCTGGGCAACAGCGAG 540
Db 148 SerLeuGluAsnGlnLeuAspGlnMetProArgArgPheTyrLeuAlaGlyAsnGlnGlu 167
QY 541 CAAGAGTCTTAAGTACCAAGCAACAAAGCAGACAAAGCAGACGAAGAAGCTTACCATAT 600
Db 168 GlnGluPheLeuLysTyrGlnGlnGln-----GlnAspArg 176
QY 601 AGCCCATACAGCCGCAAGTACGCTAGCAAGAGAGAGCGTGAATTTAGCCCTCGAGGA 660
Db 177 -----GlnGlnGlySerGlnSerGlnLysGly 186
QY 661 CAGCAGCCGCGAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 720
Db 187 LysGln-----GlnGluGluAsnGluGlySerAsnIleLeu 199
QY 721 AGCGGCTTACGCGGAGTTCCTGGAACAGCGCTTCAGGTTGACGACAGCAGATAGTG 780
Db 200 SerGlyPheAlaProGluPheLeuLysGluAlaPheGlyVal---AsnMetGlnIleVal 218
QY 781 CAAACCTTAAGCGCGAGACCGAGAGTGAAGAGAGGAGCCATTTGACAGTGAAGGGA 840
Db 219 ArgAsnLeuGlnGlyGluAsnGluGluGluAspSerGlyAlaIleValThrValLysGly 238
QY 841 GGCTCAGATCTTGACCCGAGATAGAAAGAGAGCGTCCCGACAGAGAGGAGATACGAT 900
Db 239 GlyLeuArgValThrAlaProAlaMetArgLysProGlnGlnGluAspAspAsp 258
QY 901 GAAGATGAA-----TATGAATACCATGAAGAGGATAGAGCGCTGGCAGCGGGA 948
Db 259 GluGluGlnProGlnCysValGluThrAspLysGlyCysGlnArgGlnSerLysArg 278
QY 949 AGCAGAGCGAGGGGAATGTTGTAAGAGACGATCTGACCCGCAAGTGTCTAAAGAAC 1008
Db 279 SerArg-----AsnGlyIleAspGluThrIleCysThrMetArgLeuArgGlnAsn 295
QY 1009 ATTGGTAGAACACATCCCTGATCTACAACTCAAGCTGTTGCTTCACTCAAACTGCC 1068
Db 296 IleGlyGlnAsnSerProAspIleTyrAsnProGlnAlaGlySerIleThrAla 315
QY 1069 AACCATCTCAACCTTCTAATACTTAGTGGCTTGGACCTAGTGTGAATATGGAATCTC 1128
Db 316 ThrSerLeuAspPheProAlaLeuTrpLeuLysLeuSerAlaGlnTyrGlySerLeu 335
```

Db 41 GlnCysalaGlyValAlaLeuSerArgCysThrLeuAsnArgAsnAlaLeuArgArgPro 60
Qy 181 TTCTACTCCATGCTCCCGAGGAGATCTTCATCCAGCAAGAGGGGATCTTTGGGTG 240
Db 61 SerTyrThrAsnGlyProGlnGluIleTyrArgGlnGlyAsnGlyIlePheGlyMet 80
Qy 241 ATATTCCCTGTTCTCTAGACACTATGAGAGCCTCACACAAAGTCTGCTGATCTCAG 300
Db 81 IlePheProGlyCysProSerThrTyrGlnGluProGlnGluSerGlnGlnArgGlyArg 100
Qy 301 TCCCAAGACACCAAGACAGCTCTCCAAGGAGAGACCAAGCAAGCAAGACAGATAGT 360
Db 101 SerGlnArgPro-----GlnAspArg 107
Qy 361 CACCAAGAGGTGACCGTTTCGATGAGGGTGATCTCATTCAGTTCCTCCACCGGTGTGCT 420
Db 108 HisGlnIysValHisArgPheArgGluGlyAspLeuIleAlaValProThrGlyValAla 127
Qy 421 TTCTGGCTCTACACGACACGACGACACTGATGTTGCTGCTGTTCTTCTACTGACCAAC 480
Db 128 TrpTrpMetTyrAsnAsnGluAspThrProValValAlaValSerIleIleAspThrAsn 147
Qy 481 AACCAACGACACCAAGCTTGTAGTTCCTCCAGGAGATTCAATTTGGCTGGGAACACCGAG 540
Db 148 SerLeuGluAsnGlnLeuAspGlnMetProArgPheTyrLeuAlaGlyAsnGlnGlu 167
Qy 541 CAAGAGTTCTTAAGTACCAGCAACAAAGCAGACAAAGCAGACGACGAGCAAGCTTACCATAT 600
Db 168 GlnGluPheLeuIysTyrGlnGln-----176
Qy 601 AGCCCATACAGCCCGCAAGTCCAGCTAGACAGAGCGTGAATTTAGCCCTCGAGGA 660
Db 177 -----GlnGlnGlySerGlnSerGlnLysGly 186
Qy 661 CAGCACAGCCGACAGAAACGAGCAGCAAGAAAGAAAGAGGTGGAACATCTTC 720
Db 187 LysGln-----GlnGluGluAsnGluGlySerAsnIleLeu 199
Qy 721 AGCGGCTTCACGCGGAGTTCCTGGAACAAGCCTTCAGGTTCAGCAGACAGACAGATAGTG 780
Db 200 SerGlyPheAlaProGluPheLeuLysGluAlaPheGlyVal---AsnMetGlnIleVal 218
Qy 781 CAACAACCTAGAGCGGAGCCGAGAGTGAAGAGAGGGGCCATTTGACAGTGCAGGGGA 840
Db 219 ArgAsnLeuGlnGlyGluAsnGluGluAspSerGlyAlaIleValThrValLysGly 238
Qy 841 GGCCTCAGAAATCTTGAGCCAGATAGAAAGAGAGCTGCCAGCAAGAGAGGAATACGAT 900
Db 239 GlyLeuArgValThrAlaProAlaMetArgLysProGlnGlnGluAspAspAsp 258
Qy 901 GAAGATGAA-----TATGAATACGATGAAGAGGATAGAAGCGGTGCGCAGGGGA 948
Db 259 GluGluGluGlnProGlnCysValGluThrAspLysGlyCysGlnArgGlnSerLysArg 278
Qy 949 AGCAGAGCGGGGGAATGTTATGAAGACGATCTGCACCCGCAAGTGTCTAAAGAAC 1008
Db 279 SerArg-----AsnGlyIleAspGluThrIleCysThrMetArgLeuArgGlnAsn 295
Qy 1009 ATTGGTAGAAACAGATCCCTGACATCTACAACCTCAAGCTGTTCACTCAAAAGTCC 1068
Db 296 IleGlyGlnAsnSerSerProAspIleTyrAsnProGlnAlaGlySerIleThrAla 315
Qy 1069 AACGATCTCAACCTTCTAATCTTAGGTGGCTCGACCTAGTGTGTAATATGGAATCTC 1128
Db 316 ThrSerLeuAspPheProAlaLeuTrpLeuLeuLysLeuSerAlaGlnTyrGlySerLeu 335
Qy 1129 TACAGGATGCATTTGTTGCTGCTCACTACACCAACGACACACATCATATATCGA 1188
Db 336 ArgLysAsnAlaMetPheValProHisTyrThrLeuAsnAlaAsnSerIleIleTyrAla 355
Qy 1189 TTGAGGGGCGGCTCAGCTGCAAGTCTGACAGCAACGCAACGAGTGTACGAGAG 1248
Db 356 LeuAsnGlyArgAlaLeuValGlnValValAsnCysAsnGlyGluArgValPheAspGly 375

Qy 1249 GAGCTTCAAGAGGGTCACGTGCTTGTGGTGCCACAGAACTTCGCCGTCGCTGGAAGTCC 1308
Db 376 GluLeuGlnGluGlyGlyValLeuIleValProGlnAsnPheAlaValAlaLysSer 395
Qy 1309 CAGAGCGAGAACTTCGAATACGTGGCATTCAAGACAGACTCAAGGCCACGATAGCCAAC 1368
Db 396 GlnSerAspAsnPheGluTyrValSerPheLysThrAsnAspArgProSerIleGlyAsn 415
Qy 1369 CTCGCCGTGAAAACCTCCGTTCATAGATAACCTCCGAGGAGGTGGTTGCAAAATTCATAT 1428
Db 416 LeuAlaGlyAlaAsnSerLeuLeuAlaLeuProGluGluValIleGlnHisThrPhe 435
Qy 1429 GGCTTCAAGAGGAGCAGGCAAGCAGCTTAAAGAACAAACCCCTTCAAGTCTTCGTT 1488
Db 436 AsnLeuLysSerGlnGlnAlaArgGlnValLysAsnAsnAsnPropheSerPheLeuVal 455
Qy 1489 CCACCGCTCTCAGCAGTCTCCGAGGGCTGTGGCT 1521
Db 456 ProProGlnGluSerGlnArgAlaValAla 466

Search completed: April 8, 2005, 06:49:19
Job time : 200.941 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 8, 2005, 03:43:40 ; Search time 203.93 Seconds

(without alignments)
5783.472 Million cell updates/sec

Title: US-10-728-051-3

Perfect score: 2691

Sequence: 1 cggcagcaaccggagagaa.....ctccgagggtgtgcttaa 1524

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=slp
-Q=/cgn2_1/USPTO.spool_p/US10728051/runat_07042005_125044_14041/app.query.fasta_1.4757
-DB=A_Geneseq -QFWT=fastan -SUFFIX=rag -MINMATCH=0 -1 -LOOPTCT=0 -LOOPTXT=0
-UNITS=bits -START=1 -END=1 -WATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10728051 @CGN 1.1.672 @runat_07042005_125044_14041 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_16Dec04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2673	99.3	510	6	ABU52484
2	2673	99.3	510	7	ADG27544 Peanut ar
3	2667	99.1	510	4	AAU04708 Anaphylac
4	2665	99.0	507	8	ADM12139 Arachis h
5	2663	99.0	510	2	AAI15246 Peanut al
6	2652	98.6	512	2	AAI40912 Ara h 3 a
7	2638	98.0	526	3	AAB33601 Modified
8	2638	98.0	526	4	AAU04711 Modified
9	2638	98.0	526	4	AAU05036 Modified
10	2615	97.2	507	8	ADO38357 Peanut al

11	1502	55.8	481	5	ABG71266	Glycine m
12	1502	55.8	481	7	ADH89253	G. max gl
13	1502	55.8	481	7	ADL90187	Soybean g
14	1502	55.8	481	8	ADG43988	G. max gl
15	1499.5	55.7	485	5	ABG71265	Glycine m
16	1499.5	55.7	485	7	ADH89247	G. max gl
17	1499.5	55.7	485	7	ADL90186	Soybean g
18	1499.5	55.7	485	8	ADG43982	G. max gl
19	1496	55.6	495	3	AAH80994	Soybean g
20	1496	55.6	495	3	ABG71264	Glycine m
21	1496	55.6	495	7	ADH89245	G. max gl
22	1496	55.6	495	7	ADL90168	Soybean g
23	1496	55.6	495	8	ADG43980	G. max gl
24	1496	55.6	511	7	ADL90190	Soybean g
25	1492.5	55.5	485	6	ABU52502	Soybean g
26	1492.5	55.5	485	7	ADG27563	Soybean g
27	1488	55.3	484	2	AAH40949	Soybean g
28	1466	54.5	495	4	AAE10365	Soybean g
29	1034.5	38.4	457	8	ADO43062	Cashew nu
30	995.5	37.0	484	1	AAH71081	Sequence
31	986	36.6	517	5	ABG71267	Glycine m
32	986	36.6	517	7	ADL90189	Soybean g
33	973.5	36.2	562	7	ADH89249	G. max gl
34	973.5	36.2	562	7	ADL90188	Soybean g
35	973.5	36.2	562	8	ADG43984	G. max gl
36	946.5	35.2	562	5	ABG71268	Glycine m
37	941	35.0	561	1	AAH61363	Soybean g
38	926.5	34.4	516	7	ADH89251	G. max gl
39	926.5	34.4	516	8	ADG43986	G. max gl
40	914	34.0	185	2	AAH40951	Ara h 3 a
41	901.5	33.5	516	1	AAH61362	Soybean g
42	891	33.1	291	8	ADO60333	Cholester
43	845.5	31.4	499	7	ADC08275	Rice prot
44	841	31.3	499	7	ADC08135	Rice prot
45	841	31.3	499	7	ADC08279	Rice prot

ALIGNMENTS

RESULT 1

ABU52484

ID ABU52484 standard; protein; 510 AA.

AC ABU52484;

DT 10-MAR-2003 (first entry)

XX Peanut Ara h3 protein sequence.

DE Peanut Ara h3 protein sequence.

XX Peanut; allergy; Ara h1; Ara h2; Ara h3; IgE binding site;

KW anaphylactic food allergen; anti-allergenic; vaccine; wound healing.

XX Arachis hypogaea.

OS WO200274250-A2.

XX 26-SEP-2002.

XX 18-MAR-2002; 2002WO-US009108.

XX 16-MAR-2001; 2001US-0276822P.

PR 18-MAR-2002; 2002US-00276822.

XX (PANA-) PANACEA PHARM.

XX Caplan M, Sosin H, Sampson H, Bannon GA, Burks WA, Cockrell G;

PI Compadre CM, Connaughton C, Helm RM, King NE, Kopper RA, Maleki SJ;

PI Rabjohn PA, Shain DS, Stanley JS;

XX WPI; 2003-018765/01.

DR N-PSDB; ABX70612.

XX New modified anaphylactic food allergen, useful for preventing or

PT treating allergic reactions associated with e.g. anaphylactic allergens.

XX PS Claim 27; Fig 68B; 300pp; English.

XX CC The invention relates to a modified anaphylactic food allergen has an amino acid sequence that is substantially identical to that of natural anaphylactic food allergen, except for a cysteine residue that has been modified so that it cannot participate in the disulphide bond. The modification may also comprise mutation of the IgE binding sites to reduce allergenicity. Also included are: (1) a method of making a modified anaphylactic food allergen; (2) a nucleotide molecule encoding for causing a site specific mutation in the modified anaphylactic food allergen; (3) a transgenic plant or animal expressing the modified anaphylactic food allergen; (4) a method of treating an individual by reducing the clinical response to a natural anaphylactic food allergen; and an isolated fragment of peanut allergen Ara h 1. The modified anaphylactic food allergen is useful for preventing or treating allergic reactions associated with any natural allergen such as food, insect, rubber or preferably anaphylactic allergens. It is also useful for treating wounds in mammals such as bovine, canine, feline, caprine, ovine, porcine, murine or equine species. The present sequence is a peanut allergen (e.g. Ara h1, h2 or h3)

XX Sequence 510 AA;

Alignment Scores:

Pred. No.: 3,77e-236 Length: 510
Score: 2673.00 Matches: 507
Percent Similarity: 100.00% Conserved: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.33% Indels: 0
DB: Gaps: 0

US-10-728-051-3 (1-1524) x ABUS2484 (1-510)

QY 1 CGCAGCAACCGAGAGAACGGCTGCCAGTTCACAGCGCTCAATCGCGAGAGACTGAC 60
DB 4 ArgGlnGlnProGluGluAsnAlaCysGlnPheGlnArgLeuAsnAlaGlnArgProAsp 23
QY 61 AATCGCATTAATCAGAGGCGGTTCATTGACACTTGGAAACCCCAACACAGAGAGTTC 120
DB 24 AsnArgIleGlnSerGluGlyGlyTyrIleGluThrTrpAsnProAsnGlnGluPhe 43
QY 121 GAATCGCGCGCTCGCCCTCTCTCGCTTAGTCTCCTCGCGCAACGCCCTTCGTAGCCT 180
DB 44 GluCysAlaGlyValAlaLeuSerArgLeuValLeuArgArgAsnAlaLeuArgPro 63
QY 181 TTCTACTCAATGCTCCAGGAGATCTTCATCCAGCAAGAGGGGATCTTTGGGTG 240
DB 64 PheTyrSerAsnAlaProGlnGluIlePheIleGlnGlnGlyArgGlyTyrPheGlyLeu 83
QY 241 ATATTCCCTGTTGTCCTAGACACTATGAAGAGCCTCACACAAAGTCTCGTCTCAG 300
DB 84 IlePheProGlyCysProArgHisTyrGluGluProHisThrGlnGlyArgArgSerGln 103
QY 301 TCCCAAGACCAACAGACTCTCCAGAGAGAGACAAAGCCAAACAGCAACAGATAGT 360
DB 104 SerGlnArgProProArgArgLeuGlnGlyGluAspGlnSerGlnGlnArgAspSer 123
QY 361 CACAGAGGTGACCGTTTCGATGAGGTGATCTCATTCAGTCTCCACCGGTGTGCT 420
DB 124 HisGlnLysValHisArgPheAspGluGlyAspLeuIleAlaValProThrGlyValAla 143
QY 421 TTCTGGCTCTACACGACCAACGACACTGATGTTGTTGCTGTTTCTTCTTACTGACCAAC 480
DB 144 PheTrpLeuTyrAsnAspHisAspThrAspValValAlaValSerLeuThrAspThrAsn 163
QY 481 AACACGACCAACGACTTGATTCAGTTCCTCCAGAGATTCATTTGGTGGGAACAGGAG 540
DB 164 AsnAsnAspAsnGlnLeuAspGlnPheProArgArgPheAsnLeuAlaGlyAsnThrGlu 183
QY 541 CAGAGTCTTAGTACCAGCAACAAAGCAGACAAAGCAGACGAAAGCTTACCATAT 600

DB 184 GlnGluPheLeuArgTyrGlnGlnSerArgGlnSerArgArgSerLeuProTyr 203
QY 601 AGCCCATACAGCCCAAAAGTCAGCTAGACAAAGAGCGTGAATTTAGCCCTCGAGA 660
DB 204 SerProTyrSerProGlnSerGlnProArgGlnGluArgGluPheSerProArgGly 223
QY 661 CAGCACAGCCGAGAGAACGAGCAGGACAGAGAAAGAAAGAGAGGTTGGAACATCTTC 720
DB 224 GlnHisSerArgArgGluArgAlaGlyGlnGluGluAsnGluGlyGlyAsnIlePhe 243
QY 721 AGCGCTTCACGCCGAGTCTCTGAAACAAGCCTTCCAGGTTGACGACAGACAGATAGT 780
DB 244 SerGlyPheThrProGluPheLeuGlnAlaPheGlnValAspArgGlnIleVal 263
QY 781 CAAAACCTTAAGAGCGACACGAGAGTGAAGAAGAGGAGCCATTGTGACAGTGGGGA 840
DB 264 GlnAsnLeuArgGlyGluThrGluSerGluGluGlyAlaIleValThrValArgGly 283
QY 841 GGCCTCAGAACTTTCAGCCAGATGAAGAAGACAGCTGCCGACGAGAGAGAAATACGAT 900
DB 284 GlyLeuArgIleLeuSerProAspArgIysArgArgAlaAspGluGluGluTyrAsp 303
QY 901 GAAGATGAATATGATACGATGAAGAGATGAAGCGCTGGCAGCGGAAGCAGAGCAGG 960
DB 304 GluAspGluTyrGluTyrAspGluGluAspArgArgGlyArgGlySerArgGlyArg 323
QY 961 GCGAATGTTATTGAAGAGACGATCTGCACCGCAAGTGTCTAAAGAACAATTCGTAGAAC 1020
DB 324 GlyAsnGlyIleGluGluThrIleCysThrAlaSerAlaLysLysAsnIleGlyArgAsn 343
QY 1021 AGATCCCTGACATCTACAAACCTCAAGCTGTTTCACTCAAAACTGCCCAACCATCTCAAC 1080
DB 344 ArgSerProAspIleTyrAsnProGlnAlaGlySerLeuLysThrAlaAsnAspLeuAsn 363
QY 1081 CTTCTAATACTATTAGTGTGCTTGCACCTAGTCTGAATATGGAATCTCTACAGGAATGCA 1140
DB 364 LeuLeuIleLeuArgTrpLeuGlyProSerAlaGluTyrGlyAsnLeuTyrArgAsnAla 383
QY 1141 TTGTTTGTGCTCCTACCAACCAACCAACGACACAGCATCATATATCGATTGAGGGACGG 1200
DB 384 LeuPheValAlaHisTyrAsnThrAsnAlaHisSerIleIleTyrArgLeuArgGlyArg 403
QY 1201 GCTCAGTCAAGTCTGTGGACACGACACGACAGAGTGTACGACGAGAGCTTCAAGAG 1260
DB 404 AlaHisValGlnValValAspSerAsnGlyAsnArgValTyrAspGluGluGlnGlu 423
QY 1261 GGTCACTGCTTGTGTGTCACAGAACTTCGCGCGCTCGCTGGAAGTCCCGAGCGAGAAC 1320
DB 424 GlyHisValLeuValValProGlnAsnPheAlaValAlaGlyLysSerGlnSerGluAsn 443
QY 1321 TTCGAATACGTGGCATTTCAAGACAGACTCAAGGCCCGAGCATAGCCAACTCCCGGTGAA 1380
DB 444 PheGluTyrValAlaPheLysThrAspSerArgProSerIleAlaAsnLeuAlaGlyGlu 463
QY 1381 AACTCGTCAATAGATAACCTCGCGGAGAGTGGTTCAAATTCATATGCTCCCAAGC 1440
DB 464 AsnSerValIleAspAsnLeuProGluGluValValAlaAsnSerTyrGlyLeuGlnArg 483
QY 1441 GAGCAGCGAAGCGAGCTTAAAGAACCAACACCCCTTCAAGTCTTCTGTTCCCGCTCAG 1500
DB 484 GluGlnAlaArgGlnLeuLysAsnAsnAsnProPheLysPheValProSerGln 503
QY 1501 CAGTCTCCGAGGCTGTGGCT 1521
DB 504 GlnSerProArgAlaValAla 510
RESULT 2
ADG27544
ID ADG27544 standard; protein; 510 AA.
XX AC
XX ADG27544;
XX AC
XX 26-FEB-2004 (first entry)
DT

XX Peanut allergen Ara h3.
 XX
 XX
 KW Peanut; plant; allergen; Ara h1; Ara h2; Ara h3; glycinin A2B1a; Jug n1;
 KW antiallergic; vulnerary; anaphylactic food allergen; IgE; allergy; wound.
 XX
 XX Arachis hypogaea.
 XX
 XX
 PN US2003020980-A1.
 XX
 XX 30-OCT-2003.
 XX
 XX 18-MAR-2002; 2002US-00100303.
 XX
 XX 29-DEC-1995; 95US-0009455P.
 PR 23-SEP-1996; 96US-00717933.
 PR 31-JAN-1998; 98US-0073283P.
 PR 13-FEB-1998; 98US-0074590P.
 PR 13-FEB-1998; 98US-0074624P.
 PR 13-FEB-1998; 98US-0074633P.
 PR 29-JUN-1998; 98US-00106872.
 PR 27-AUG-1998; 98US-00141220.
 PR 13-NOV-1998; 98US-00191593.
 PR 29-JAN-1999; 99US-00240557.
 PR 29-JAN-1999; 99US-00241101.
 PR 11-FEB-1999; 99US-00248673.
 PR 11-FEB-1999; 99US-00248674.
 PR 02-MAR-1999; 99US-0122450P.
 PR 02-MAR-1999; 99US-0122452P.
 PR 02-MAR-1999; 99US-0122560P.
 PR 02-MAR-1999; 99US-0122565P.
 PR 02-MAR-1999; 99US-0122566P.
 PR 11-MAR-1999; 99US-00267719.
 PR 28-JAN-2000; 2000US-00494096.
 PR 16-MAR-2001; 2001US-0276822P.
 XX
 XX (CAPL/) CAPLAN M J.
 PA (SOSI/) SOSIN H B.
 PA (SAMP/) SAMPSON H.
 PA (BANN/) BANNON G A.
 PA (BURK/) BURKS A W.
 PA (COCK/) COCKRELL G.
 PA (COMP/) COMPADRE C M.
 PA (CONN/) CONNAUGHTON C.
 PA (HELM/) HELM R M.
 PA (KING/) KING N E.
 PA (KOPP/) KOPPER R A.
 PA (MALE/) MALEKI S J.
 PA (RABJ/) RABJOHN P A.
 PA (SHIN/) SHIN D S.
 PA (STAN/) STANLEY J S.
 XX
 XX Caplan MJ, Sosin HB, Sampson H, Bannon GA, Burks AW, Cockrell G;
 PI Compadre CM, Connaughton C, Helm RM, King NE, Kopper RA, Maleki SJ;
 PI Rabjohn PA, Shin DS, Stanley JS;
 XX
 DR WPI; 2003-875632/81.
 DR N-PSDB; ADG27543.
 XX
 XX
 PT New modified anaphylactic food allergen comprising a cysteine residue
 PT which has been modified so that it cannot participate in the disulfide
 PT bond, useful for treating allergic reactions or wounds.
 XX
 XX Example 16; SEQ ID NO 89; 194pp; English.
 XX
 CC The invention relates to a modified anaphylactic food allergen whose
 CC amino acid sequence is substantially identical to that of a natural
 CC anaphylactic food allergen. The natural anaphylactic food allergen
 CC includes at least one cysteine residue that participates in a disulphide
 CC bond when the natural anaphylactic food allergen is in its native
 CC conformation, except that the cysteine residue has been modified so that
 CC it cannot participate in the disulphide bond. Also included are a method
 CC of making a modified anaphylactic food allergen, a nucleotide molecule

CC encoding a modified anaphylactic food allergen defined above, a
 CC nucleotide molecule for causing a site specific mutation in a gene
 CC encoding a natural anaphylactic food allergen, a transgenic plant or
 CC animal expressing a modified anaphylactic food allergen defined above, a
 CC method of treating an individual by reducing the clinical response to a
 CC natural anaphylactic food allergen by administering a modified
 CC anaphylactic food allergen and an isolated fragment of peanut allergen
 CC Ara h 1, comprising at least 10 consecutive amino acids of ADG27464 or
 CC ADG27465. About 10-17% of the amino acids have been modified in at least
 CC one IGE epitope or all the IGE epitopes recognised when the natural
 CC anaphylactic food allergen is contacted with serum IgE from individual(s)
 CC allergic to the natural anaphylactic food allergen. The invention
 CC discloses Peanut allergens Ara h1, Ara h2, Ara h3 (and their encoding
 CC cDNAs), Soybean Glycinin A2B1a and IGE-binding epitopes of the English
 CC walnut allergen Jug n1. The modified anaphylactic food allergen can be
 CC used for treating allergic reactions or wounds. The present sequence
 CC represents a Peanut allergen of the invention (or its fragment).
 XX
 XX SQ Sequence 510 AA;
 Alignment Scores:
 Pred. No.: 3,77e-236 Length: 510
 Score: 2673.00 Matches: 507
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.33% Indels: 0
 DB: 7 Gaps: 0
 US-10-728-051-3 (1-1524) x ADG27544 (1-510)
 QY 1 CGGCACGACCGGAGGAGAACGGTCCAGTTCACGGCGCTCAATGGCGAGACCTGAC 60
 DB 4 ArgGlnGlnProGluGluAsnAlaCysGlnPheGlnArgLeuAsnAlaGlnArgProAsp 23
 QY 61 AATCGCATTGAATCAGAGGCGGTTTACATTGACACTTGAACCCCAACCAACAGGAGTTC 120
 DB 24 AsnArgIleGluSerGluGlyTyrlleGluThrTrpAsnProAsnGlnGluPhe 43
 QY 121 GAATGCGCGGGTCCGCCCTCTCTCGCTTAGTCTCTCCGGCGCAACGCCCTTCGTAGSCCT 180
 DB 44 GluCysAlaGlyValAlaLeuSerArgLeuValLeuArgArgAsnAlaLeuArgArgPro 63
 QY 181 TTCTACTCCAAATGCTCCCGAGGAGATCTTCATCCAGCAAGGAGGGGATCTTTGGTTG 240
 DB 64 PheTyrSerAsnAlaProGlnGluIlePheIleGlnGlnGlyArgGlyTyrPheGlyLeu 83
 QY 241 ATATTCCTCGTTGTCTTAGACACTATGAGAGCCCTCACACAAAGTCTCGTATCTCAG 300
 DB 84 IlePheProGlyCysProArgHisTyrGluGluProHisThrGlnGlyArgArgSerGln 103
 QY 301 TCCCAAGACCAACCAAGAGCTCTCCAGGAGAGAGACCAAGCCAAAGCAGCAACGAGATAGT 360
 DB 104 SerGlnArgProProArgArgLeuGlnGlyGluAspGlnSerGlnGlnArgAspSer 123
 QY 361 CACCAAGAGTGCACCGTTTCGATGAGGTGATCTCATTCGACGTTCCCGGTGTGCT 420
 DB 124 HisGlnLysValHisArgPheAspGluGlyAspLeuIleAlaValProThrGlyValAla 143
 QY 421 TTCTGCTCTACACACCAACCAACGACACACTGATGTTGTGTTGTTCTCTTACTGACACCAAC 480
 DB 144 PheTrpLeuTyrAsnAspHisAspThrAspValValAlaValSerLeuThrAspThrAsn 163
 QY 481 AACACGACCAACCAAGCTTGCATGTTCCCGAGAGATTCATATTTGGTGGGAACACGAG 540
 DB 164 AsnAsnAspAsnGlnLeuAspGlnPheProArgArgPheAsnLeuAlaGlyAsnThrGlu 183
 QY 541 CAAAGATCTTAAAGTACCAAGCAACAAAGCAGCAAAAGCAGAGCAAGAGCTTACCATAT 600
 DB 184 GlnGluPheLeuArgTyrGlnGlnGlnSerArgGlnSerArgArgSerLeuProTyr 203
 QY 601 AGCCCATACAGCCCGCAAGTACGCTAGCAGAGAGAGCGTCAATTTAGCCCTCCAGGA 660
 DB 204 SerProTyrSerProGlnSerGlnProArgGlnGluArgGluPheSerProArgGly 223

Db 64 PheTyrSerAsnAlaProGlnGluIlePheIleGlnGlyArgGlyTyrPheGlyLeu 83
Qy 241 ATATTCCCTGGTGTCTAGACACTATCAAGAGCCTCACACAAAGTGTGTCGATCTCAG 300
Db 84 IlePheProGlyCysProArgHisTyrGluGluProHisThrGlnGlyArgSerGln 103
Qy 301 TCCCAAGACCACCAAGAGCTCTCAAGAGGAGAGACCAAGCCACCAAGACGAGTAGT 360
Db 104 SerGlnArgProProArgArgLeuGlnGlyGluAspGlnSerGlnGlnArgAspSer 123
Qy 361 CACCAGAGGTGACCGCTTTCAGTACGAGGTGATCTCATTCAGTTCCTCCACCGGTGTGCT 420
Db 124 HisGlnLysValHisArgPheAspGluGlyAspLeuIleAlaValProThrGlyValAla 143
Qy 421 TTCTGGCTCTACACGACACGACACTGATGTTGTTGCTGTTCTTCTTACTGACCAAC 480
Db 144 PheTrpLeuTyrAsnAspHisAspThrAspValValAlaValSerLeuThrAspThrAsn 163
Qy 481 AACACACACACACAGCTTGATCAGTTCCTCCAGGAGATTCATTTGGCTGGGAACACGAG 540
Db 164 AsnAsnAspAsnGlnLeuAspGlnPheProArgArgPheAsnLeuAlaGlyAsnThrGlu 183
Qy 541 CAAGAGTCTTAAAGTACCAGCAACAAAGCAGACAAAGCAGACGACGACGACGACATAT 600
Db 184 GlnGluPheLeuArgTyrGlnGlnGlnSerArgGlnSerArgArgSerLeuProTyr 203
Qy 601 AGCCATACAGCCCGCAAGTACGCTTAGACAAGAGCGTGAATTTAGCCCTCGAGGA 660
Db 204 SerProTyrSerProGlnSerGlnProArgGlnGluArgGluPheSerProArgGly 223
Qy 661 CAGCACGCGCAGAGAACGAGCAGGACAGAGAGAAAGAAAGAGAGTGGAAACATCTTC 720
Db 224 GlnHisSerArgArgGluArgAlaGlyGlnGluGluAsnGluGlyGlyAsnIlePhe 243
Qy 721 AGCGGCTTCAGCGCGGAGTCTCTGGAACAGCCCTTCAGTTCAGCAGACAGACATAGTG 780
Db 244 SerGlyPheThrProGluPheLeuGluGlnAlaPheGlnValAspArgGlnIleVal 263
Qy 781 CAACACCTAAGAGCGAGACCGAGAGTGAAGAGAGGAGGACCATTTGACAGTGGAGGA 840
Db 264 GlnAsnLeuArgGlyGluThrGluSerGluGluGlyAlaIleValThrValArgGly 283
Qy 841 GGCTCAGATCTTTCAGCCAGATAGAAAGAGAGCTGCCAGCAGAAAGAGGATACGAT 900
Db 284 GlyLeuArgIleLeuSerProAspArgLysArgArgAlaAspGluGluGluTyrAsp 303
Qy 901 GAGATCAATATGAATACGATGAGAGGATAGAGGCGTCGAGGGAAGCAGAGGCGAG 960
Db 304 GluAspGluTyrGluTyrAspGluGluAspArgArgGlyArgGlySerArgGlyThr 323
Qy 961 GGGAAATGGTATTGAAGAGACGATCTGCCCGCAAGTCTTAAAGAAACATTTGGTAGAAAC 1020
Db 324 GlyAsnGlyIleGluGluThrIleCysThrAlaSerAlaLysLysAsnIleGlyArgAsn 343
Qy 1021 AGATCCCTCAGATCTTCAACCCCTCAAGCTGTTCTCACTCAAACTGCCAACGATCTCAAC 1080
Db 344 ArgSerProAspIleTyrAsnProGlnAlaGlySerLeuLysThrAlaAsnAspLeuAsn 363
Qy 1081 CTTCTAATACTAGTGGCTGGACCTAGTGTGCTGAATATGGAATCTTACAGGAATGCA 1140
Db 364 LeuLeuIleLeuArgTrpLeuGlyProSerAlaGluTyrGlyAsnLeuTyrArgAsnAla 383
Qy 1141 TTGTTTTCGCTCACTACACCAACCAAGCAGCAGCATCATATATCATTTGAGGGGACGG 1200
Db 384 LeuPheValAlaHisTyrAsnThrAsnAlaHisSerIleIleTyrArgLeuArgGlyArg 403
Qy 1201 GCTCAGTGAAGTCGTGGACAGCAACGCGCAACAGAGTGTACGACGAGGAGCTTCAAGAG 1260
Db 404 AlaHisValGlnValAspSerAsnGlyAsnArgValTyrAspGluGluLeuGlnGlu 423
Qy 1261 GGTACAGTGTGTTGGTGGCCACAGAACTTGCCTGCTGGAAAGTCCAGAGGAGAAC 1320

Db 424 GlyHisValLeuValValProGlnAsnPheAlaValAlaGlyLysSerGlnSerGluAsn 443
Qy 1321 TTCCAATACGTGGCAATTCAGACAGACTCAAGCCCAAGCATAGCAACCTCGCCGGTCAA 1380
Db 444 PheGluTyrValAlaPheLysThrAspSerArgProSerIleAlaAsnLeuAlaGlyGlu 463
Qy 1381 AACTCCGTCTAGATACCTGCGGAGGAGGTGTTGCAAAATTCATATGSCCTCCAAAGG 1440
Db 464 AsnSerValIleAspAsnLeuProGluGluValAlaAsnSerTyrGlyLeuGlnArg 483
Qy 1441 GAGCAGCAAGGAGCGCTTAAGAAACAACCCCTTCAAGTTCCTCGTTCCACCCGCTCAG 1500
Db 484 GluGlnAlaArgGlnLeuLysAsnAsnAsnProPheLysPheValProProSerGln 503
Qy 1501 CAGTCTCCGAGGCTGTGGCT 1521
Db 504 GlnSerProArgAlaValAla 510
RESULT 4
ADM12139
ID ADM12139 standard; protein; 507 AA.
XX
AC ADM12139;
XX
DT 20-MAY-2004 (first entry)
XX
DE Arachis hypogaea 3 (Ara h3) protein.
XX
KW antigen presentation enhancing hybrid polypeptide; mammalian II-Key;
KW MHC class II; antibacterial; virucide; fungicide; antirheumatic;
KW antiarthritic; neuroprotective; dermatological; immunosuppressive;
KW antiinflammatory; antidiabetic; antithyroid; immune;
KW rheumatoid arthritis; multiple sclerosis; lupus erythematosus;
KW diabetes mellitus; myasthenia gravis; autoimmune thyroiditis;
KW scleroderma; dermatomyositis; pemphigus.
XX
OS Arachis hypogaea.
XX
PN US2003235594-A1.
XX
PD 25-DEC-2003.
XX
PF 17-SEP-2002; 2002US-00245871.
XX
PR 14-SEP-1999; 99US-00396813.
PR 17-JUL-2002; 2002US-00197000.
XX
PA (ANTI-) ANTIGEN EXPRESS INC.
XX
PI Humphreys R, Xu M;
XX
PS WPI; 2004-070554/07.
XX
CC Novel II-Key/antigen presentation enhancing hybrid polypeptide, useful
PT for treating infections, rheumatoid arthritis, multiple sclerosis, lupus
PT erythematosus and diabetes mellitus.
XX
PS Example 3; Page 21; 87pp; English.
XX
CC The invention relates to a novel antigen presentation enhancing hybrid
CC polypeptide. The novel polypeptide has an N-terminal element consisting
CC of 4-16 residues of a mammalian II-Key peptide and its non-N-terminal
CC deletion modifications, a chemical structure covalently linking the N-
CC terminal element to an MHC class II-presented epitope of a C-terminal
CC element. The C-terminal element comprises an antigenic epitope, which
CC binds to an antigenic peptide binding site of an MHC class II molecule.
CC The antigen presentation enhancing hybrid polypeptide has the following
CC activities: antibacterial, virucide, fungicide, antirheumatic,
CC antiarthritic, neuroprotective, dermatological, immunosuppressive,
CC antiinflammatory, antidiabetic, and antithyroid. The antigen presentation
CC enhancing hybrid polypeptide is useful for modulating the immune response
CC in an individual and for treating infections (such as bacteria, virus,
CC parasite and fungus), rheumatoid arthritis, multiple sclerosis, lupus

CC erythematous, diabetes mellitus, myasthenia gravis, autoimmune
 CC thyroiditis, scleroderma, dermatomyositis and pemphigus. This sequence
 CC represents a mammalian II key related protein of the invention.

XX
 SQ Sequence 507 AA;

Alignment Scores:

Pred. No.: 2,04e-235 Length: 507
 Score: 2665.00 Matches: 506
 Percent Similarity: 99.80% Conservative: 0
 Best Local Similarity: 99.80% Mismatches: 1
 Query Match: 99.03% Indels: 0
 DB: 8 Gaps: 0

US-10-728-051-3 (1-1524) x ADM12139 (1-507)

QY	1	CGGACACACCGGAGGAGACCGTCCAGTTCACGCGCTCAATCGGAGACCTGAC	60
DB	1	ArgGlnProGluGluAsnAlaCysGlnPheGlnArgLeuAsnAlaGlnArgProAsp	20
QY	61	AATCGCATTAATCAGAGGCGGTACATTGAGACTTGGAAACCCCAACACGAGGAGTTC	120
DB	21	AsnArgIleGluSerGluGlyTyrIleGluThrTrpAsnProAsnAsnGlnGluPhe	40
QY	121	GAATGCGCGGCTCGGCTCTCTCGCTTAGTCTCTCGGCGCAACGCGCTTCGTAGCCT	180
DB	41	GluCysAlaGlyValAlaLeuSerArgLeuValLeuArgArgAsnAlaLeuArgPro	60
QY	181	TTCTACTCCATGCTCCGAGGAGACTTCATCCAGCAGGAGGAGTACTTGGGTG	240
DB	61	PheTyrSerAsnAlaProGlnGluIlePheIleGlnGlyArgGlyTyrPheGlyLeu	80
QY	241	ATATTCCCTGGTGTCTGACACTATGAAGAGCCTCACACAAAGTGTGCTCATCTCAG	300
DB	81	IlePheProGlyCysProArgHisTyrGluGluProHisThrGlnGlyArgSerGln	100
QY	301	TCCCAAGACCAACAGACTCTCCAGGAGAGACCAAGCAACAGCAACGAGATAGT	360
DB	101	SerGlnArgProArgArgLeuGlnGlyGluAspGlnSerGlnGlnArgAspSer	120
QY	361	CACGAGAGTGCACCGTTCGATGAGGTGATCTCATTCAGTCTCCACCGGTGTGCT	420
DB	121	HisGlnLysValHisArgThrAspGluGlyAspLeuIleAlaValProThrGlyValAla	140
QY	421	TTCTGGCTTACCAACACGACACTGATGTTGTTGCTGTTTCTCTTACTGACCAAC	480
DB	141	PheTrpLeuTyrAsnAspHisAspThrAspValAlaValSerLeuThrAspThrAsn	160
QY	481	AACAACGACCAACGAGTTGATCAGTTCCTCCAGGAGATTCATTTGGCTGGGAAACGGAG	540
DB	161	AsnAsnAspAsnGlnLeuAspGlnPheProArgPheAsnLeuAlaGlyAsnThrGlu	180
QY	541	CAAGAGTTCTTAAGGTACACGACCAACAGCAGACCAAGCAGACGAGAGCTTACCATAT	600
DB	181	GlnGluPheLeuArgTyrGlnGlnGlnSerArgGlnSerArgArgSerLeuProTyr	200
QY	601	AGCCCATACACCGCGAAAGTTCAGCTCAGACAGAGAGCGTGAATTTAGCCCTCGAGGA	660
DB	201	SerProTyrSerProGlnSerGlnProArgGlnGluGluArgGluPheSerProArgGly	220
QY	661	CAGCAGCGCGAGAGAACGAGGAGGACAGAGAGAGAAACGAAGTGGAAACATCTTC	720
DB	221	GlnHisSerArgArgGluArgAlaGlyGlnGluGluAsnGluGlyGlyAsnIlePhe	240
QY	721	AGCGGTTTCACGCGGAGTTCCTGGACAAAGCTTCAGGTTCCAGCAGACAGATAGTG	780
DB	241	SerGlyPheThrProGluPheLeuGluGlnAlaPheGlnValAspAspArgGlnIleVal	260
QY	781	CAAAACCTTAAGAGCGGAGACCGAGAGTGAAGAGAGGAGGAGCCATTGTGACAGTGAGGGA	840
DB	261	GlnAsnLeuArgGlyGluThrGluSerGluGluGluGlyAlaIleValThrValArgGly	280
QY	841	GGCCTCAGATCTTTGAGCCCGCATAGAAAGAGAGCTGCCAGCAGAGAGAGATACGAT	900

DB	281	GlyLeuArgIleLeuSerProAspArgLysArgAlaAspGluGluGluGlyArgP	300
QY	901	GAAAGATGAATATGAATACCATCAAGAGGATGAAGCGTGGCAGGAGGAGCAGG	960
DB	301	GluAspGluTyrGluTyrAspGluGluAspArgArgGlyArgGlySerArgGlyArg	320
QY	961	GGGAATGGTATTGAAGAGACGATCTGCACCGCAAGTGTAAAAAGAACATTGGTGAAC	1020
DB	321	GlyAsnGlyIleGluGluThrIleCysThrAlaSerAlaLysLysAsnIleGlyArgAsn	340
QY	1021	AGATCCCTGACATCTACAAACCTCAAGCTGTTTCACTCAAACTGCCAACCATCTCAAC	1080
DB	341	ArgSerProAspIleTyrAsnProGlnAlaGlySerLeuLysThrAlaAsnAspLeuAsn	360
QY	1081	CTTCTAATTAATTAAGTGGCTTCGACCTAGTGTGTAATATGAAATCTTACAGGAATGCA	1140
DB	361	LeuLeuIleLeuArgTyrLeuGlyProSerAlaGluTyrGlyAsnLeuTyrArgAsnAla	380
QY	1141	TTGTTTGTGCTCCTACCAACCAACGACACAGCATCATATATCGATTGAGGAGCAGG	1200
DB	381	LeuPheValAlaHisTyrAsnThrAsnAlaHisSerIleIleTyrArgLeuArgGlyArg	400
QY	1201	GCTCAGTCAAGTCTGTCGACACGACGACGACGAGTGTACGAGGAGGCTTCAAGAG	1260
DB	401	AlaHisValGlnValValAspSerAsnGlyAsnArgValTyrAspGluGluGlnGlu	420
QY	1261	GCTCAGTCTGCTGTGTCGACAGAACTTCGCGCGTCTGGAAGTCCGAGAGCGAGAAC	1320
DB	421	GlyHisValLeuValValProGlnAsnPheAlaValAlaGlySerGlnSerGluAsn	440
QY	1321	TTTCGAATACGTGGCATTTCAAGACAGACTCAAGGCGGACGATAGCCAACTCGCGGTGAA	1380
DB	441	PheGluTyrValAlaPheLysThrAspSerArgProSerIleAlaAsnLeuAlaGlyGlu	460
QY	1381	AACTCCGTATAGATAACCTCGCGGAGGAGTGTTCGAAATTCATATGCGCTCCAAAGG	1440
DB	461	AsnSerValIleAspAsnLeuProGluGluValAlaAsnSerTyrGlyLeuGlnArg	480
QY	1441	GAGCAGCGCAGGAGCGCTTAAGAACCAACACCCCTTCAAGTCTTCTGTTCCACCGTCTCAG	1500
DB	481	GluGlnAlaArgGlnLeuLysAsnAsnAsnProPheLysPheValProProSerGln	500
QY	1501	CAGTCTCCGAGGCGTGTGGCT	1521
DB	501	GlnSerProArgAlaValAla	507

RESULT 5

AA15246
 ID AA15246 standard; protein; 510 AA.

XX

AC AA15246;

XX

DT 17-OCT-2003 (revised)

DT 09-NOV-1999 (first entry)

XX

DE Peanut allergen, Ara h 3, amino acid sequence.

XX

XX allergy; immune response; transgenic; allergen; epitope;

KW immunoglobulin E; Ig E; binding site; peanut.

XX

OS Arachis hypogaea.

XX

PN WO9938978-A1.

XX

XX 05-AUG-1999.

XX

PF 29-JAN-1999; 99WO-US002031.

XX

PR 31-JAN-1998; 98US-0073283P.

PR

PR 13-FEB-1998; 98US-0074590P.

PR

PR 13-FEB-1998; 98US-0074624P.

PR

PR 13-FEB-1998; 98US-0074633P.

AC AAY40912;
 DT 17-OCT-2003 (revised)
 DT 06-DEC-1999 (first entry)
 XX Ara h 3 allergen sequence.
 XX Peanut; allergen; Ara h 1; IgE; immunoglobulin E; epitope; Ara h 3;
 XX allergic reaction; ss.
 XX Arachis hypogaea.
 XX
 PH Key Location/Qualifiers
 FT Region 187..188
 FT /note= "these two amino acid residues are not indicated
 FT in the Ara h 3 sequence provided in the sequence listing
 FT (seq ID No: 6) and Fig 51 of the specification"
 FT Misc-difference 374
 FT /note= "encoded by CCT"
 XX
 PN WO9945961-A1.
 XX
 XX 16-SEP-1999. 99WO-US005494.
 XX
 XX 12-MAR-1999; 99WO-US005494.
 XX
 XX 12-MAR-1998; 98US-0077763P.
 XX
 XX 11-MAR-1999; 99US-00077763.
 XX
 XX (UYAR-) UNIV ARKANSAS.
 XX
 XX Burks W, Helm RM, Cockrell G, Bannon GA, Stanley JS, Shin DS;
 PI Sampson H, Compadre CM, Huang SK, Maleki SJ, Kopper RA;
 XX
 XX WPI; 1999-551218/46.
 DR N-PSDB; AAZ22280.
 XX
 XX Tertiary structure of peanut allergen Ara h 1 for protection of a host
 PT animal from allergic reaction.
 XX
 PS Claim 8; Fig 11; 193pp; English.
 CC
 CC The invention provides a tertiary structure for the peanut allergen Ara H
 CC 1. The Ara H 1 allergen is found to contain 23 linear IgE-binding
 CC epitopes. The invention also provides an isolated recombinant peanut
 CC allergen designated Ara h 3 and a nucleotide molecule encoding the peanut
 CC allergen Ara h 3. Molecules of the invention are used to protect a host
 CC animal from allergic reaction, particularly using a modified allergen
 CC which is less reactive with IgE. The invention may also be used to ensure
 CC that the allergen is not introduced into genetically modified food. The
 CC present sequence represents a DNA encoding the Ara h 3 allergen. (Updated
 CC on 17-OCT-2003 to standardise OS field)
 XX
 SQ Sequence 512 AA;
 Alignment Scores:
 Pred. No.: 3,21e-234 Length: 512
 Score: 2652.00 Matches: 506
 Percent Similarity: 99.41% Conservative: 0
 Best Local Similarity: 99.41% Mismatches: 1
 Query Match: 98.55% Indels: 2
 DB: 2 Gaps: 1
 US-10-728-051-3 (1-1524) x AAY40912 (1-512)
 QY 1 CGGCAGCAACCGGAGGAGACGGTCCAGTTCAGCGCTCAATCGCAGAGACCTGAC 60
 DB 4 ArgGlnGlnProGluGluAsnAlaCysGlnPheGlnArgLeuAsnAlaGlnArgProAsp 23
 QY 61 AATCGATTGAATCAGAGGCGGTTACATTGAGACTTGGAACTTGGAAACCCCAACACGAGGATTC 120
 DB 24 AsnArgileGluSerGluGlyGlyTyrlleGluThrTyrAsnProAsnAsnGlnGluPhe 43

QY 121 GAATGCGCGGGTCCGCTCTCTCGCTTAGTCTCTCCGCCGCAACGCCCTTCGTAGCCT 180
 DB 44 GluCysAlaGlyValAlaLeuSerArgLeuValLeuArgArgAsnAlaLeuArgArgPro 63
 QY 181 TTCTACTCCAATGCTCCCGAGGAGATCTTCATCCAGCAAGAGGAGATCTTTGGGTTG 240
 DB 64 PheTyrSerAsnAlaProGlnGluIlePheIleGlnGlnGlyArgGlyTyrPheGlyLeu 83
 QY 241 ATATTCCCTGGTTGCTCTAGACACTATGAAGAGCCTCACACACAAGTCGTCTCATCTCAG 300
 DB 84 IlePheProGlyCysProArgHisTyrGluGluProHisThrGlnGlyArgArgSerGln 103
 QY 301 TCCAAAGACCAACAGACGCTCTCCAAAGGAGAGACCAAGGCAACAGCAACAGATAGT 360
 DB 104 SerGlnArgProArgArgLeuGlnGlyGluAspGlnSerGlnGlnGlnArgAspSer 123
 QY 361 CACCAGAGGTCCACCGTTTCGATGAGGAGTATCTCATTCGAGTTCCTCCACCGGTGTGCT 420
 DB 124 HisGlnLysValHisArgPheAspGluGlyAspLeuIleAlaValProThrGlyValAla 143
 QY 421 TTCTGCTCTACAAACGACCACTGATGTTGTTGTTGTTGTTCTTCTTACTGACACCAAC 480
 DB 144 PheTrpLeuTyrAsnAspHisAspThrAspValValAlaValSerLeuThrAspThrAsn 163
 QY 481 AACAAACGACACGCTTCATCAGTTCCTCCAGGAGATTCAATTTGGCTGGGAACACGAG 540
 DB 164 AsnAsnAspAsnGlnLeuAspGlnPheProArgArgPheAsnLeuAlaGlyAsnThrGlu 183
 QY 541 CAA-----GAGTTCTTAAGTTACCGCAACCAAGCAGACCAAGCAGACGAGAACTTA 594
 DB 184 GlnGluPheGluPheLeuArgTyrGlnGlnGlnSerArgGlnSerArgArgArgSerLeu 203
 QY 595 CCATATAGCCCATACAGCCCGCAAGTACGCTAGACAAAGAGAGCGTCAATTTAGCCCT 654
 DB 204 ProTyrSerProTyrSerProGlnSerGlnProArgGlnGluGluArgGluPheSerPro 223
 QY 655 CGAGGACACACAGCCGCGCAGAGAACGAGCAGACAAAGAGAGAAAGAAACCAAGGTGAAAC 714
 DB 224 ArgGlyGlnHisSerArgArgGluArgAlaGlyGlnGluGluGluAsnGluGlyGlyAsn 243
 QY 715 ATCTTCAGCGGCTTCACGCCGGAGTTCCTGGAAACAGCCTCCAGGTTCCACGACGAG 774
 DB 244 IlePheSerGlyPheThrProGluPheLeuGluGlnAlaPheGlnValAspAspArgGln 263
 QY 775 ATAGTGCAGAACTTAAGAGCGCAGAGCCGAGAGTGAAGAGAGAGGAGCCATTGTGACAGTG 834
 DB 264 IleValGlnAsnLeuArgGlyGluThrGluSerGluGluGluGlyAlaIleValThrVal 283
 QY 835 AGGGAGGCTCAGAACTCTTGAGCCCGCAGATGAAGAGAGACGTGCCGACCAAGAGAGGAA 894
 DB 284 ArgGlyGlyLeuArgileLeuSerProAspArgLysArgAlaAspGluGluGlu 303
 QY 895 TAGGATGAAGATGAATATGAATACGATGAAGAGGATAGAGCGCTGGCAGGGGAGCAG 954
 DB 304 TyrAspGluAspGluTyrGluTyrAspGluGluAspArgArgGlyArgGlySerArg 323
 QY 955 GGCAGGGGGAATGGTATTGAAGAGACGATCTCGACCGCAAGTGCTAAAAGAACATTGGT 1014
 DB 324 GlyArgGlyAsnGlyIleGluGluThrIleCysThrAlaSerAlaIleGlyAsnIleGly 343
 QY 1015 AGAAACAGATCCCTGACATCTACACCCCTCAAGCTGGTTCTCACTCAAACTGCCAACGAT 1074
 DB 344 ArgAsnArgSerProAspIleTyrAsnProGlnAlaGlySerLeuLysThrAlaAsnAsp 363
 QY 1075 CTCACCTCTTAATCTTAGTGGCTTGGACCTAGTGTGTAATATGGAATCTCTACAGG 1134
 DB 364 LeuAsnLeuLeuIleLeuArgTrpLeuGlyLeuSerAlaGluTyrGlyAsnLeuTyrArg 383
 QY 1135 AATGCAATTGTTGTCGCTCACTTACCAACCAACGACGACGATCATATATCGATTGAGG 1194
 DB 384 AsnAlaLeuPheValAlaHisTyrAsnThrAsnAlaHisSerIleIleTyrArgLeuArg 403
 QY 1195 GGACGGGCTCAGTCAAGTCGTGACACGCAACGCGACAGAGTGTACGACGAGGCTT 1254

Db 404 GlyArgAlaHisValGlnValValAspSerAsnGlyAsnArgValTyAspGluGluLeu 423
 QY 1255 CAAGAGGTCACGTGTTGGTGCCACAGAACTTCGCCCTCGCTGAAAGTCCACAGAGC 1314
 Db 424 GlnGluGlyHisValLeuValProGlnAsnPheAlaValAlaGlyLysSerGlnSer 443
 QY 1315 GAGAACTTCGAATACGTGGCATTCAGACAGACTCAAGGCCAGCATAGCCACCTCGCC 1374
 Db 444 GluAsnPheGluTyValAlaPheLysThrAspSerArgProSerIleAlaAsnLeuAla 463
 QY 1375 GGTGAAAACTCCGTCATAGATAACCTCGCCGAGGAGGTGGTGCATAATTCATATGGCCTC 1434
 Db 464 GlyGluAsnSerValIleAspAsnLeuProGluValValAlaAsnSerTyGlyLeu 483
 QY 1435 CAAGAGGAGCAGCAAGCGCTTAAGAACCAACACCCCTTCAGTTCTTCGTTCCACCG 1494
 Db 484 GlnArgGluGlnAlaArgGlnLeuLysAsnAsnProPheLysPheValProPro 503
 QY 1495 TCTCAGCAGTCTCGAGGGCTGGCT 1521
 Db 504 SerGlnGlnSerProArgAlaValAla 512

RESULT 7 AAB33601

ID AAB33601 standard; protein; 526 AA.

XX AAB33601;
 AC
 XX 12-SEP-2003 (revised)
 DT 22-JAN-2001 (first entry)
 XX
 DE Modified Ara h 3 amino acid sequence.
 XX
 KW Allergy; allergic reaction; allergen; anaphylactic antigen; peanut;
 KW Ara h 1; Ara h 2; Ara h 3; epitope; binding; immunoglobulin E; IGE;
 KW food antigen; sensitising; immune response; anti-allergic.
 XX

OS Arachis hypogaea.

XX WO200051647-A2.

XX 08-SEP-2000.

XX 03-MAR-2000; 2000WO-US005655.

XX 03-MAR-1999; 99US-0122960P.

PR 06-DEC-1999; 99US-00455294.

XX (MOUN) MOUNT SINAI SCHOOL MEDICINE.

XX Sampson HA;

PI WPI; 2000-611341/58.

DR WPI; 2000-611341/58.

XX Non-human animal sensitized to an antigen, useful as an animal model for
 PT studying allergic reactions to allergens, such as those in food and in
 PT the environment.

XX Example 6; Fig 17C; 124pp; English.

CC The present invention describes an animal model which can be used for
 CC studying allergic reactions to allergens. The animal is sensitized to a
 CC selected antigen by administering the antigen itself or a nucleic acid
 CC encoding the antigen, where preferably the antigen is an anaphylactic
 CC antigen. The sensitized animal can then be used to screen for compounds
 CC which help to prevent, ameliorate, or cure allergic conditions in
 CC humans. The animal model can be used for studying allergic reactions to
 CC allergens, such as those in food (peanuts, fruit, berry, nut, bean, milk,
 CC dairy products), or in the environment (weed pollen, grass pollen, tree
 CC pollen, mite, animal, animal dander, fungal, and insect antigens).
 CC AAB33478 to AAB33601 represent sequences which are used in examples from
 CC the present invention to specifically examine the peanut allergy, and the

CC peanut anaphylactic antigens Ara h 1 to 3. (Updated on 12-SEP-2003 to
 CC standardise OS field)
 XX
 SQ Sequence 526 AA;
 Alignment Scores:
 Pred. No.: 6,27e-233 Length: 526
 Score: 2638.00 Matches: 502
 Percent Similarity: 99.01% Conservative: 0
 Best Local Similarity: 99.01% Mismatches: 5
 Query Match: 98.03% Indels: 0
 DB: 3 Gaps: 0

US-10-728-051-3 (1-1524) x AAB33601 (1-526)

QY 1 CGGCACAAACCGAGGAGAAACGGTGCCTCCAGCGCTCAATGCGCAGACCTGAC 60
 Db 5 ArgGlnGlnProGluGluAsnAlaCysGlnPheGlnArgLeuAsnAlaGlnArgProAsp 24
 QY 61 AATCGCATTCGAATCAGAGGGCGGTTCATTTGAGACTTGGAAACCCCAACACCGAGGTTC 120
 Db 25 AsnArgIleGluSerGluGlyGlyTyIleGluThrTrpAsnAlaAsnGlnGluPhe 44
 QY 121 GAATGCGCGGCGTCCGCTCTCTCGTTAGTCTCTCGCGCGCAACGCCCTTCGTAGGCCT 180
 Db 45 GluCysAlaGlyValAlaLeuSerArgLeuValLeuArgAsnAlaLeuArgArgPro 64
 QY 181 TTCTACTCCAATGCTCCCGAGGAGATCTTCATCCAGCAAGAAAGGGGATCTTTGGGTG 240
 Db 65 PheTy-SerAsnAlaProGlnGluIlePheIleGlnGlnGlyArgGlyTyPheGlyLeu 84
 QY 241 ATATTCCTCGTGTCTCTAGACACTATGAGAGCCTCACACACAGTCTCGATCTCAG 300
 Db 85 IlePheProGlyCysProArgHisTyGluGluProHisThrGlnGlyArgArgSerGln 104
 QY 301 TCCCAAGACCAACAGAGCGTCCCAAGGAGAAAGACCAAGCCCAACAGCAACGAGTAGT 360
 Db 105 SerGlnArgProArgArgLeuGlnGlyGluAspGlnSerGlnGlnArgAspSer 124
 QY 361 CACCAGAGGTGCACCGTTCGATGAGGGTGATCTCATTCAGTTCCTCCCGGTGTGCT 420
 Db 125 HisGlnLysValHisArgPheAspGluGlyAspLeuIleAlaValProThrGlyValAla 144
 QY 421 TTCTGCTCTACAAACGACACGACACTGATGTTGTTGCTGTTCTTCTTACTGACACCAAC 480
 Db 145 PheTrpLeuTyTrAsnAspHisAspThrAspValValAlaValSerLeuThrAspThrAsn 164
 QY 481 AACACGACAAACCGAGTTCGATCAGTTCCCGAGGAGATTCATTTGGTGGGAAACACGAG 540
 Db 165 AsnAsnAspAsnGlnLeuAspGlnPheProArgArgPheAsnLeuAlaGlyAsnThrGlu 184
 QY 541 CAAGAGTTCCTTAAGGTACCAACGACAAAGCAGACAAAGCAGACGAAAGAGCTTACCATAT 600
 Db 185 GlnGluPheLeuArgTyGlnGlnGlnSerArgGlnSerArgArgSerLeuProTy 204
 QY 601 AGCCCATACAGCCCGCAAGTCAAGCTAGCCTAGACAGAGAGCGTGAATTTAGCCCTCAGGA 660
 Db 205 SerProTySerProGlnSerGlnProArgGlnGluGluArgGluPheSerProArgGly 224
 QY 661 CAGCAGCGCGCAGAGAACGAGCAGACAGAGAAAGAAACGAAAGTGGAAACATCTTC 720
 Db 225 GlnHisSerArgArgGluArgAlaGlyGlnGluGluGluGluGluGlyAsnIlePhe 244
 QY 721 AGCGGCTTCACGCGCGGAGTTCCTGGAAACCAAGCCCTTCAGGTTGACGACAGACAGATAGT 780
 Db 245 SerGlyPheThrProGluAlaLeuPheGlnAlaPheGlnValAspAspArgGlnIleVal 264
 QY 781 CAAAACCTAAGAGCGGAGCCGAGAGTGAAGAGAGAGGAGGCCATTGTGACAGTGGAGGA 840
 Db 265 GlnAsnLeuArgGlyGluThrGluSerGluGluGluGlyAlaIleValThrValArgGly 284
 QY 841 GGCCTCAGATCTTGAGCCCGCAGATAGAAAGCAGAGCTGCCCGCAAGAGAGGAGATACGAT 900

Db 285 GlyLeuArgAlaLeuSerProAspArgLysArgAlaAspGluGluGluIuTyArg 304
QY 901 GAAGATGAATATCAATACGATGAAGAGGATAGAGCGTGCAGGGAAGCAGAGGAGG 960
Db 305 GluAspGluTyAlaTyArgGluGluAspArgArgGlyArgGlySerArgGlyArg 324
QY 961 GGAAATGGTATTGAAGAGAGCATCTGCACCGCAAGTGCTAAAAAGAACATTTGGTAGAAC 1020
Db 325 GlyAsnGlyIleGluGluThrIleCysThrAlaSerAlaLysLysAsnIleGlyArgAsn 344
QY 1021 AGATCCCTCAGATCTACACCTCTCAAGCTGGTTCACCTCAAACTGCCAACGATCTCAAC 1080
Db 345 ArgSerProAspIleTyArgProGlnAlaGlySerLeuLysThrAlaAsnAspLeuAsn 364
QY 1081 CTTCTAATAGTGGTGGCTTGGACCTAGTGTGCTGAATATGCAATCTTACAGCAATGCA 1140
Db 365 LeuLeuIleLeuArgTrpLeuGlyProSerAlaGluTyArgLysLeuTyArgAsnAla 384
QY 1141 TTGTTTGTGCTCACTACACCAACCAACGACACAGCATCATATATCGATTGAGGAGCGG 1200
Db 385 LeuPheValAlaHisTyArgThrAsnAlaHisSerIleIleTyArgLeuArgGlyArg 404
QY 1201 GCTCAGTGCAGTCTGGACAGCAACGCAACAGAGTGTACACGAGGAGCTTCAAGG 1260
Db 405 AlaHisValGlnValValAspSerAsnGlyAsnArgValTyArgGluLeuGlnGlu 424
QY 1261 GGTCAAGTGTGCTGTGTCACAGAACTTCGCGCTGCTGGAAGTCCACAGAGCAGAAC 1320
Db 425 GlyHisValLeuValProGlnAsnPheAlaValAlaGlyLysSerGlnSerGluAsn 444
QY 1321 TTGCAATAGTGCATTCAAGACAGACTCAAGCCGACGATACGCCAATCTGCGGTGAA 1380
Db 445 PheGluTyValAlaPheLysThrAspSerArgProSerIleAlaAsnLeuAlaGlyGlu 464
QY 1381 AACTCCGTATAGTACCTGCGGAGGAGTGTGCAATTCATATGCTTCAAGG 1440
Db 465 AsnSerValIleAspAsnLeuProGluGluValValAlaAsnSerTyArgLeuGlnArg 484
QY 1441 GAGCAGCAGGAGGAGCTTAAAGAACAAACACCCCTTCAAGTTCCTGCTCCACGCTCTCAG 1500
Db 485 GluGlnAlaArgGlnLeuLysAsnAsnProPheLysPhePheValProProSerGln 504
QY 1501 CAGTCTCCGAGGCTGTGGCT 1521
Db 505 GlnSerProArgAlaValAla 511

RESULT 8
AAU04711
ID AAU04711 standard; protein; 526 AA.
XX
AC AAU04711;
XX
XX 23-OCT-2001 (first entry)
XX
DE Modified anaphylactic antigen Ara h 3.
XX
XX Ara h 3; anaphylactic antigen; immunoglobulin E; IgE; immunogenic;
KW allergy; mast cell; basophil; mouse.
XX
XX Mus sp.
OS Synthetic.
XX
XX W0200140264-A2.
XX
XX 07-JUN-2001.
XX
XX 06-DEC-2000; 2000WO-US033124.
XX
XX 06-DEC-1999; 99US-00455294.
PR 23-JUN-2000; 2000US-0213765P.
PR 27-SEP-2000; 2000US-0235797P.
XX
XX (PANA-) PANACEA PHARM LLC.

(UYAR-) UNIV ARKANSAS.
(MOUN) MOUNT SINAI SCHOOL MEDICINE NEW YORK.

Bannon CA, Burks WA, Caplan MJ, Sampson H, Sosin H;
WPI; 2001-381378/40.

Antigenic fragments useful for reducing anaphylactic risk and reducing
the severity and/or number of allergic symptoms in individuals sensitive
to antigens, have reduced ability to bind immunoglobulin E.

Disclosure; Fig 12; 100pp; English.

The sequence represents the amino acid sequence of modified anaphylactic
antigen Ara h 3, which has been altered to disrupt immunoglobulin E (IgE)
binding sites. Ara h 3 is an anaphylactic antigen (A), which was used to
design antigenic peptides having a reduced ability to bind IgE as
compared with the intact (A), or having a sequence substantially
identical to a portion of sequence of an antigen that includes at least
one IgE binding site, where at least one IgE binding site of the peptide
is altered. The antigenic peptides are used in a composition which is
useful for reducing risk or severity of allergic reaction to an antigen.
This is done by identifying an individual at risk of allergic reaction to
an antigen by identifying prior display of allergic symptoms when exposed
to the antigen, or a familial relationship with an individual who
previously displayed allergic symptoms when exposed to the antigen.
Following this an antigen-specific IgE present on one or more mast cells
or basophils in the individual's serum is identified. The individual is
then contacted with a peptide corresponding to a portion of the antigen,
which is selected, formulated, and delivered so that binding of the
peptide to antigen-specific IgE is reduced as compared with IgE binding
of intact antigen. The composition is also useful for treating and
preventing allergic reactions

SQ Sequence 526 AA;

Alignment Scores:

Pred. No.:	6,27e-233	Length:	526
Score:	2638.00	Matches:	502
Percent Similarity:	99.01%	Conservative:	0
Best Local Similarity:	99.01%	Mismatches:	5
Query Match:	98.03%	Indels:	0
DB:	4	Gaps:	0

US-10-728-051-3 (1-1524) x AAU04711 (1-526)

QY 1 CGGACGACACCGAGGAGAACCGTGCAGGCTCCAGCGCTCAATGCGCAGAGCCTGAC 60

Db 5 ArgGlnGlnProGluGluAsnAlaCysGlnPheGlnArgLeuAsnAlaGlnArgProAsp 24

QY 61 AATCGCATTGAATCAGAGGCGGTTACATTGAGACTTGGAAACCCCAACACGAGGTTTC 120

Db 25 AsnArgIleGluSerGluGlyGlyTyrlleGluThrTrpAsnAlaAsnGlnGluPhe 44

QY 121 GAATGCGCGGCTGCGCTCTCTCGTTAGTCTCCGCGCGCAACGCCCTTCGTAGGCCT 180

Db 45 GluCysAlaGlyValAlaLeuSerArgLeuValLeuArgAsnAlaLeuArgArgPro 64

QY 181 TTCTACTCCAAATGCTCCCGAGGAGATCTTCATCCAGAGGAGGAGTACTTTGGGTG 240

Db 65 PheTySerAsnAlaProGlnGlnIlePheIleGlnGlnGlyArgGlyTyrlleGlyLeu 84

QY 241 ATATTCCCTGGTTGTCTAGACACTATGAAGAGCCTTCACACAAAGGTTCGTCTCAG 300

Db 85 IlePheProGlyCysProArgHisTyrlleGluGluProHisThrGlnGlyArgSerGln 104

QY 301 TCCCAAGACACCAAGACCTCTCCAGAGAGACCAAGCAACAGCAACAGAGTAGT 360

Db 105 SerGlnArgProProArgArgLeuGlnGlyGluAspGlnSerGlnGlnGlnArgAspSer 124

QY 361 CACCAGAAGGTGCACGTTTTCGATGAGGTGTATCTCATTGCAGTTCCCAACCGGTGTGCT 420

Db 125 HisGlnLysValHisArgPheAspGluGlyAspLeuIleAlaValProThrGlyValAla 144

Qy	421	TTCTGGCTCTACAACGACCGACACTGATGTTGTTGCTGTTCTCTTACTGACACCAAC	480
Db	145	PheTrpLeuTyAenAspHisAspThrAspValValAlaValSerLeuThrAspThrAsn	164
Qy	481	AACAACGACAAACAGCTTGATCAGTCTCCCGAGGAGATTCAAATTTGGCTGGGACACCGGAG	540
Db	165	AsnAsnAspAsnGlnLeuAspGlnPheProArgPheAsnLeuAlaGlyAsnThrGlu	184
Qy	541	CAAGAGTTCTTAAGGTACCAGCAACAAGACAGACAAAGACAGACGAAGAAGCTTACCATAT	600
Db	185	GlnGluPheLeuArgTyGlnGlnGlnSerArgGlnSerArgArgAspSerLeuProTyR	204
Qy	601	AGCCCATACAGCCCGCAAAAGTCACGCTAGACAAGAAGCGTGAAATTTAGCCCTCGAGGA	660
Db	205	SerProTySerProGlnSerGlnProArgGlnGluArgGluPheSerProArgGly	224
Qy	661	CAGCACAGCCGACAGAAACGACGAGGACAAAGAAGAAAACGAAGTGTGAAAACATCTTC	720
Db	225	GlnHisSerArgArgGluArgAlaGlyGlnGluGluAsnGluGlyGlyAsnIlePhe	244
Qy	721	AGCGGTTACGCCGGAGTTCCTGGAAACAAGCCTTCCAGGTTGACGACAGATAGTG	780
Db	245	SerGlyPheThrProGluAlaLeuPheGlnAlaPheGlnValAspAspArgGlnIleVal	264
Qy	781	CAAAAACCTTAAGAGCGGACCGACGAGTGAAGAAGAGGAGGCCATTGTGACAGTGAGGGGA	840
Db	265	GlnAsnLeuArgGlyThrGluSerGluGluGluGlyAlaIleValThrValArgGly	284
Qy	841	GGCCTCAGAATCTTGAGCCCAAGATAGAAAGAGAGCTGCCGACGAAGAAGAGGAATACGAT	900
Db	285	GlyLeuArgAlaLeuSerProAspArgLysArgAlaAspGluGluGluGluTyRasp	304
Qy	901	GAAGATGAATATGAATACGATGAAGAGGATAGAAGCGTCGCGGGGAAACGACAGGCAGG	960
Db	305	GluAspGluTyAlaTyAspGluGluAspArgArgGlyArgGlySerArgGlyArg	324
Qy	961	GGGAATGGTATTGAAGAGACGATCTGCACCGCAGTCTTAAAGAGACATTTGTTAGAAAC	1020
Db	325	GlyAsnGlyIleGluGluThrIleCysThrAlaSerAlaLysLysAsnIleGlyArgAsn	344
Qy	1021	AGATCCCTCGACATCTCAACCCCTCAAGCTGGTTCACTCAAAAACGTCGAACGATCTCAAC	1080
Db	345	ArgSerProAspIleTyAenProGlnAlaGlySerLeuLysThrAlaAsnAspLeuAsn	364
Qy	1081	CTTCTAATACATTAGTGCTGGACCTAGTGTCTGAATGTAATGAAATCTCTACAGGAATGCA	1140
Db	365	LeuLeuIleLeuArgTyrLeuGlyProSerAlaGluTyArgLysAsnLeuTyArgAsnAla	384
Qy	1141	TTGTTTCTGCTGCTACTCAACACCAACGACACAGCATCATATATCATTTGAGGGGACGG	1200
Db	385	LeuPheValAlaHisTyAsnThrAsnAlaHisSerIleIleTyArgLeuArgGlyArg	404
Qy	1201	GCTCACGTGCAAGTCGTGGACAGCAACGGCAACAGAGTGTACGACGAGGAGCTTCAAGAG	1260
Db	405	AlaHisValGlnValValAspSerAsnGlyAsnArgValTyAspGluGluLeuGlnGlu	424
Qy	1261	GGTCACGTGCTGTGTGGCCACAGAACTTCGCCCTGCTGGAAGTGTCCACAGAGCGAGAAC	1320
Db	425	GlyHisValLeuValValProGlnAsnPheAlaValAlaGlyLysSerGlnSerGluAsn	444
Qy	1321	TTCCGAATACGTGGCATTCACACACAGACTCAAGGCCCGACGATAGCAACCTCGCGGTAA	1380
Db	445	PheGluTyRValAlaPheLysThrAspSerArgProSerIleAlaAsnLeuAlaGlyGlu	464
Qy	1381	AACTCCGTGCATAGATAAACCCTCCCGAGGAGGTGGTTGCAAAATTCATATGGCTCCAAAGG	1440
Db	465	AsnSerValIleAspAsnLeuProGluValValAlaAsnSerTyGlyLeuGlnArg	484
Qy	1441	GACGAGGCAAGCGAGCTTAAGACAACAACCCCTTCAGTCTTCTGTTCCACCGCTCTCAG	1500
Db	485	GluGlnAlaArgGlnLeuLysAsnAsnAsnProPheLysPhePheValProProSerGln	504

Qy	1501	CAGTCTCCGAGGGCTGTGGCT	1521
Db	505	GlnSerProArgAlaValAla	511
RESULT	9		
AAU05036			
ID	AAU05036	standard; protein; 526 AA.	
XX	AC	AAU05036;	
XX	XX		
DT	24-OCT-2001	(first entry)	
XX	XX		
DE	Modified anaphylactic peanut antigen Ara h 3.		
XX	XX		
KW	Anaphylactic antigen; Ara h 3; peanut; desensitisation; antigen; allergy; immunotherapy; rush immunotherapy; anaphylaxis; asthma; food allergen; milk allergen; shellfish allergen; latex; drug; environmental allergen; grass pollen; ovalbumin; insect venom; peanut oil; immunoglobulin E; IgE.		
OS	Arachis sp.		
OS	Synthetic.		
XX	XX		
FN	WO200139799-A2.		
XX	XX		
PD	07-JUN-2001.		
XX	XX		
PF	06-DEC-2000; 2000WO-US033125.		
XX	XX		
PR	06-DEC-1999; 99US-00455294.		
PR	23-JUN-2000; 2000US-0213765P.		
PR	27-SEP-2000; 2000US-0235797P.		
XX	XX		
PA	(PANA-) PANACEA PHARM LLC.		
XX	XX		
FI	Caplan M;		
PI	WPI; 2001-408258/43.		
DR	XX		
XX	XX		
PT	Preventing allergic response against antigens, e.g. food and environmental allergens such as peanut allergen or grass pollen, latex or drug, comprises administering agent, e.g. a peptide, that blocks antigen binding sites on offending IgE.		
PT	XX		
PS	Disclosure; Fig 2; 76pp; English.		
XX	XX		
CC	The sequence represents the amino acid sequence of modified anaphylactic peanut antigen Ara h 3. The protein is used for active or passive desensitisation of an individual to an antigen; for alleviating or preventing allergic reactions and for decreasing the risk of allergic reactions during immunotherapy or rush immunotherapy, anaphylaxis and asthma. The antigen may be a food allergen (e.g. peanut or milk allergen), shellfish allergen, environmental allergen (e.g. grass pollen or tree pollen), latex, drug, pollen, ovalbumin, an insect venom antigen or predominantly linear epitopes. The protein is useful for protecting an individual against subsequent inadvertent or intentional exposure to antigen, e.g. receiving blocking agent before eating a chocolate bar which may inadvertently contain peanut components or before eating foods prepared using peanut oil. Administration of the blocking agents does not result in cross-linking of anti-antigenic immunoglobulin E (IgE). After exposure to the agent, the individual's antigen sensitivity is at least temporarily reduced. Only those IgE molecules that bind the offending antigen and contribute to the risk of an allergic response are blocked		
XX	XX		
SQ	Sequence 526 AA;		
Alignment Scores:			
Pred. No.:	6,27e-233	Length:	526
Score:	2638.00	Matches:	502
Percent Similarity:	99.01%	Conservative:	0
Best Local Similarity:	99.01%	Mismatches:	5
Query Match:	98.03%	Indels:	0
DB:	4	Gaps:	0

US-10-728-051-3 (1-1524) x AAU05036 (1-526)	
QY	1 CGGACGACCGAGAGAGAACGGTCCAGTTCCAGCGCCTCAATGCGGAGAGACCTGAC 60
Db	1 LeuLeuIleLeuAaGtPLeuGlyProSerAlaGluTyrGlyAsnLeuTyrArgAsnAla 384
QY	5 ArgGlnGlnProGluGluAsnAlaCysGlnPheGlnArgLeuAsnAlaGlnArgProAsp 24
Db	5 LeuLeuIleLeuAaGtPLeuGlyProSerAlaGluTyrGlyAsnLeuTyrArgAsnAla 384
QY	61 AATCGCATTAATACAGAGCGGCTTACATTGAGACTTGGAAACCCCAACACAGAGATTC 120
Db	61 LeuPheValAlaHisTyrAsnThrAsnAlaHisSerIleIleTyrArgLeuArgGlyArg 404
QY	25 AsnArgIleIleSerGluGlyTyrIleGluThrTyrPheAsnAlaAsnAsnGlnGluPhe 44
QY	121 GAATGCGCGGCTCGCCCTCTCTCGCTTAGTCTCTCCGCGCAAGCCCTTCGTAGGCT 180
Db	121 GCTCAGTCGTCAGTCGTGAGAGCAACGCGCAACAGAGTGTACGACGAGAGCTTCAAG 1260
QY	45 GluCysAlaGlyValAlaLeuSerArgLeuValLeuArgArgAsnAlaLeuArgPro 64
Db	45 AlaHisValGlnValValAspSerAsnGlyAsnArgValTyrAspGluGluLeuGlnGlu 424
QY	181 TTCTACTCCATGCTCCAGAGAGATCTTCATCCACAGAGAGGAGGATCTTCGGTTG 240
Db	181 GGTCACTGCGTGTGTCGCCACAGAACTTCGCCGTCGCTCGCTGAAAAGTCCCAGAGCGAGAAC 1320
QY	65 PheTyrSerAsnAlaProGlnGluIlePheIleGlnGlnGlyArgGlyTyrPheGlyLeu 84
Db	65 GlyHisValLeuValValProGlnAsnPheAlaValAlaGlyLysSerGlnSerGluAsn 444
QY	241 ATATTCCCTGTTGCTCTAGACACTATGAAGAGCCTCACACAAAGCTCGTCGATCTCAG 300
Db	241 TTCGAAATACGTCGCTTCAAGACAGACTCAAGAGCCACAGAGCATAGCCACCTCGCGGTCAA 1380
QY	85 IlePheProGlyCysProArgHisTyrGluGluProHisThrGlnGlyArgSerGln 104
Db	85 PheGluTyrValAlaPheLysThrAspSerArgProSerIleAlaAsnLeuAlaGlyGlu 464
QY	301 TCCCAAGACCCACAGAGCTCTCAAGAGAGAGACAAAGCCCAACAGCAACGAGATAGT 360
Db	301 AACTCCGTCATAGATAACTGCGGAGGAGGTGGTGTGCAAAATTCATATGGCTCCAAAGG 1440
QY	105 SerGlnArgProArgArgLeuGlnGlyGluAspGlnSerGlnGlnArgAspSer 124
Db	105 AsnSerValIleAspAsnLeuProGluGluValValAlaAsnSerTyrGlyLeuGlnArg 484
QY	361 CACAGAGGTGCACCGTTTCGATGAGGTGATCTCATTCAGTTCGCCACCGGTGTGCT 420
Db	361 GAGCAGGCAAGCGACTTAAAGAACAAACCCCTTCAAGTTCCTCGTTCACCCCTCTCAG 1500
QY	125 HisGlnLysValHisArgPheAspGluGlyAspLeuIleAlaValProThrGlyValAla 144
Db	125 GluGlnAlaArgGlnLeuLysAsnAsnProPheLysPheValProProSerGln 504
QY	421 TTCTGGCTCTACACACACGACACTGATGTTGTTGCTGTTTCTCTTACTGACCAAC 480
Db	421 CAGTCTCCGAGGCTGTGGCT 1521
QY	145 PheTrpLeuTyrAsnAspHisAspThrAspValValAlaValSerLeuThrAspThrAsn 164
Db	145 GlnSerProArgAlaValAla 511
QY	481 AACACGACACCGAGTTGATGATTCCTCCAGAGAGATTCATTTGCTGGGAACAGGAG 540
Db	481 AACTCCGTCATAGATAACTGCGGAGGAGGTGGTGTGCAAAATTCATATGGCTCCAAAGG 1440
QY	165 AsnAsnAspAsnGlnLeuAspGlnPheProArgArgPheAsnLeuAlaGlyAsnThrGlu 184
QY	541 CAAGAGTTCTTAAGTACCAGCAACAAAGCAGACAAAGCAGACAGAGAGCTTACCATAT 600
Db	541 GluGlnPheLeuArgTyrGlnGlnGlnSerArgGlnSerArgArgSerLeuProTyr 204
QY	601 AGCCCATACCGCCGAAAGTACGCTTAGACAGAGAGCGTGAATTTAGCCCTCAGGGA 660
Db	601 SerProTyrSerProGlnSerGlnProArgGlnGluGluArgGluPheSerProArgGly 224
QY	661 CAGCAGACCGCCAGAGACGAGCAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
Db	661 GlnHisSerArgArgGluAaGalaGlyGlnGluGluGluGluGluGlyAsnIlePhe 244
QY	721 AGCGGCTTCACGCGGAGTTCCTCGAAACAAGCCTTCAGAGTTCACGACAGACAGATAGT 780
Db	721 Arachis hypogaea.
QY	245 SerGlyPheThrProGluAlaLeuPheGlnAlaPheGlnValAspAspArgGlnIleVal 264
Db	245 US2004058881-A1.
QY	781 CAAACCTTAAGAGCGAGACCGAGAGTGAAGAGAGGAGCCATTGTGACAGTACGAGGA 840
Db	781 25-MAR-2004.
QY	265 GlnAsnLeuArgGlyGluThrGluSerGluGluGlyAlaIleValThrValArgGly 284
QY	841 GGCCTCAGATCTTTCGCCAGATAGAAAGAGAGCTGCCGACGAAGAGAGGAATACGAT 900
Db	841 24-SEP-2002; 2002US-00253286.
QY	285 GlyLeuArgAlaLeuSerProAspArgLysArgArgAlaAspGluGluGluGluTyrAsp 304
QY	901 CAAGATGAATATGAATACGATGAAGAGATAGAAGCGGTGGCAGGGAAGCAGAGGAGG 960
Db	901 24-SEP-2002; 2002US-00253286.
QY	305 GluAspGluTyrAlaTyrAspGluGluAspArgArgGlyArgGlySerArgGlyArg 324
Db	305 (ANTI-) ANTIGEN EXPRESS INC.
QY	961 GGGATGTATGAGAGACGATCTGCACCCAGTCTGTAAGAGAGATGCTTAAAGACATTTGTTAGAAC 1020
Db	961 Humphreys RE, Xu M;
QY	325 GlyAsnGlyIleGluGluThrIleCysThrAlaSerAlaLysLysAsnIleGlyArgAsn 344
Db	325 WPI; 2004-294259/27.
QY	1021 AGATCCCTGACATCTCAACCTCAAGCTGGTTCACTCAAACTGCCCAACGATCTCAAC 1080
Db	1021 New non-naturally occurring protein or polypeptide modified by recombinant DNA techniques, useful for treating multiple sclerosis,
QY	345 ArgSerProAspIleTyrAsnProGlnAlaGlySerLeuLysThrAlaAsnAspLeuAsn 364
Db	

PT diabetes mellitus, myasthenia gravis, scleroderma, allergic rhinitis,
PT colitis, cancer or psoriasis.

XX Example 3; Page 21-22; 90pp; English.

XX The invention relates to a non-naturally occurring protein or polypeptide
CC (I) modified by recombinant DNA techniques comprising: a C-terminal
CC element comprising an MHC Class II-presented epitope; an N-terminal
CC element comprising an Ii-key motif; and an intervening element comprising
CC a sequence of 4-11 amino acid residues where the modification by
CC recombinant DNA techniques taking place within elements (b) and (c). Also
CC described are methods for: suppressing or enhancing an immune response
CC directed toward an MHC (major histocompatibility complex) Class II-
CC presented epitope of interest. Suppressing an immune response directed
CC toward an MHC Class II-presented epitope of interest comprises: providing
CC a nucleic acid sequence encoding the MHC Class II-presented epitope of
CC interest, the nucleic acid sequence encoding an Ii-key motif located 4-11
CC amino acids upstream from the N-terminal residue of the MHC Class II-
CC presented epitope of interest; and modifying the Ii-key motif to decrease
CC its conformance to the archetypal Ii-key regulatory motif. Enhancing an
CC immune response directed toward an MHC Class II-presented epitope of
CC interest comprises: providing a nucleic acid sequence encoding the MHC
CC Class II-presented epitope of interest, the nucleic acid sequence lacking
CC an Ii-key motif located 4-11 amino acids upstream from the N-terminal
CC residue of the MHC Class II-presented epitope of interest; and modifying
CC the nucleic acid sequence to introduce an Ii-key motif appropriately
CC spaced from the MHC Class II-presented epitope. The protein or
CC polypeptide of interest corresponds to a protein or polypeptide encoded
CC by an infectious pathogen selected from anthrax, EBOLA, HIV or influenza,
CC preferably vaccinia virus. The non-naturally occurring protein or
CC polypeptide (I) modified by recombinant DNA techniques is useful for
CC treating infectious diseases caused or associated with infection by a
CC bacterium, virus, parasite, fungus, rickettsia or other infectious
CC agents. It is also useful for treating rheumatoid arthritis, multiple
CC sclerosis, lupus erythematosus, diabetes mellitus, myasthenia gravis,
CC autoimmune thyroiditis, scleroderma, dermatomyositis, pemphigus, asthma,
CC allergic rhinitis, topical dermatitis, colitis, cancer, psoriasis or
CC adenomas. The present sequence represents the amino acid sequence of
CC peanut allergen Ara h 3 used in the invention.

XX Sequence 507 AA;

Alignment Scores:

Pred. No.: 8e-231 Length: 507
Score: 2615.00 Matches: 498
Percent Similarity: 98.22% Conservative: 0
Best Local Similarity: 98.22% Mismatches: 9
Query Match: 97.18% Indels: 0
DB: 8 Gaps: 0

US-10-728-051-3 (1-1524) x ADO38357 (1-507)

QY 1 CGGCAGCAACGGAGGAGAACGGTGCAGTTCAGCGCCTCAATGCGAGACCTGAC 60
DB 1 ArgGlnGlnProGluGluAsnAlaCysGlnPheGlyArgLeuAsnAlaGlnArgProAsp 20
QY 61 AATCGCATTAATCAGAGGGGGGTTCATTGAGACTTGGAAACCCCAACACCGAGGATTC 120
DB 21 AsnArgileGluSerGluGlyGlyTyrlleGluThrTrpAsnProAsnAsnGlyGluPhe 40
QY 121 GAATGCGCGGGTGCCTCTCTCGCTTAGCTTACCTCCGCGCGCAACGGCCCTTCGTAGGCCT 180
DB 41 GluCysAlaGlyValAlaLeuSerArgLeuValLeuArgArgAsnAlaLeuArgArgPro 60
QY 181 TTCTACTCCATGCTCCCGAGGAGATCTTCATCCAGCAAGGAGGGGATCTTTGGGTG 240
DB 61 PheTyrSerAsnAlaProGlnGluIlePheIleGlyGlyArgGlyTyrPheGlyLeu 80
QY 241 ATATTTCCTGTTCTCTAGACACTATGAAGAGCCTCACACAAAGGTTCGTTCATCTCAG 300
DB 81 IlePheProGlyCysProArgHisTyrGluGluProHisThrGlyGlyArgArgSerGln 100
QY 301 TCCCAAGACCAACCAAGACGTCTCCAAAGGAGAGACCAAGCCCAACAGCAACGAGATG 360

DB 101 SerGlnArgProProArgArgLeuGlnGlyGluAspGlnSerGlnGlnGlnArgAspSer 120
QY 361 CACCAGAAGGTGCACCGTTCGATGAGGGTGATCTCATTCAGTTCACCGCGGTGTGCT 420
DB 121 HisGlnLysValHisArgPheAspGluGlyPheLeuIleAlaValProThrGlyValAla 140
QY 421 TTCTGGCTTACCAACGACACGACACTGATGTTGTTGCTTCTTCTTCTTACTGACCAAC 480
DB 141 PheTrpLeuTyrAsnAspHisAspThrAspValAlaValSerLeuThrAspThrAsn 160
QY 481 AACCAACGACCAACCGTTCATCAGTTCCTCCAGAGATTCATTTGGCTGGGAAACACGAG 540
DB 161 AsnAsnAspAsnGlnLeuAspGlnPheProArgPheAsnLeuAlaGlyAsnThrGlu 180
QY 541 CAAGAGTTCCTTAAGGTACCAGCAACAAAGCAGACAAAGCAGACGAAAGCTTACCATAT 600
DB 181 GlnGluPheLeuArgTyrGlnGlnSerArgGlnSerArgArgSerLeuProTyr 200
QY 601 AGCCCATACAGCCCGCAAGTCAAGCTAGCAAGAGAGCGTGAATTTAGCCCTCGAGGA 660
DB 201 SerProTyrSerProGlnSerGlnProArgGlnGluGluArgGluPheSerProArgGly 220
QY 661 CACACAGCGCGCAGAGAACGAGCAGCAGCAGAGAGAAACCAAGTGGAAACATCTTC 720
DB 221 GlnHisSerArgArgGluArgAlaGlyGlnGluGluAsnGluGlyGlyAsnIlePhe 240
QY 721 AGCGGCTTCACGCGCGAGTTCCTGGAAACAGCCCTTCAGGTTTACGACACAGACATAGTG 780
DB 241 SerGlyPheThrProGluPheLeuGluAlaPheGlnValAspAspArgGlnIleVal 260
QY 781 CAAAACTAAGAGCGCAGACCGAGAGAGTGAAGAGAGAGGAGCCATTGTGACGAGGAGA 840
DB 261 GlnAsnLeuArgGlyGluThrSerGluSerGluGluGlyAlaIleValThrValArgGly 280
QY 841 GGCCTCAGATCTTGAGCCCGAGTACAGAGACGCTGCCACCAAGAGAGGATACGAT 900
DB 281 GlyLeuArgileLeuSerProAspArgLysArgArgAlaAspGluGluGluTyrAsp 300
QY 901 GAAGATGAATATCAATACGATGAAGAGGAGTAGAGGCGTGGCAGGGGAAACGAGGCGAGG 960
DB 301 GluAspGluTyrGluTyrAspGluGluAspArgArgArgGlyArgGlySerArgGlyArg 320
QY 961 GGAATGGTATTGAAGAGAGAGTCTGCACCGCAAGTGTCTAAAAGAACATTGTTAGAAAC 1020
DB 321 GlyAsnGlyIleGluGluThrIleCysThrAlaSerAlaLysLysAsnIleGlyArgAsn 340
QY 1021 AGATCCCTTCACATCTACACCCCTCAAGCTGCTCACTCAAACTGCCAACGATCTCAAC 1080
DB 341 ArgSerProAspIleTyrAsnProGlnAlaGlySerLeuLysThrAlaAsnAspLeuAsn 360
QY 1081 CTTCTAATACCTTAGTGGCTTGGACCTAGTGTCTGAATATCGAAATCTCTACAGGAATGCA 1140
DB 361 LeuLeuIleLeuArgTrpLeuGlyProSerAlaGluTyrGlyAsnLeuTyrArgAsnAla 380
QY 1141 TTCTTTCTCGCTCCTACCAACCAACGACACAGCATCATATATTCGATTGAGGGGACGG 1200
DB 381 LeuPheValAlaHisTyrAsnThrAsnAlaHisSerIleIleTyrArgLeuArgGlyArg 400
QY 1201 GCTCAGTGCAGTCTGGACACCAACGACAGAGTGTACGACGAGGAGCTTCAAGAG 1260
DB 401 AlaHisValGlnValValAspSerAsnGlyAsnArgValTyrAspGluGluLeuGlnGlu 420
QY 1261 GGTCACTGCTGCTGCTGGCCACAGAACTTCGCGCTCGCTGCAAGTCCGACGAGCGAGAAC 1320
DB 421 GlyHisValLeuValValProGlnAsnPheAlaValAlaGlyLysSerGlnSerGluAsn 440
QY 1321 TTCGAATACGTGGCATTTCAGACAGACTCAAGGCCAGCATAGCCACCTCGCGGTGAA 1380
DB 441 PheGluTyrValAlaPheLysThrAspSerArgProSerIleAlaAsnLeuAlaGlyGlu 460
QY 1381 AACTCCGTCTAGATTAACCTGCGGAGGAGGTGGTTCGCAATTCATATTCGCTCCAAAGG 1440

```

Db      461 AsnSerValIleAspAsnLeuProGluGluValValAlaAsnSerTyrGlyLeuGlnArg 480
QY      1441 GAGCAGGCAAGGAGCTTAAAGAACAAACCCCTTCAAGTTCCTGTTCCACCGTCTCAG 1500
Db      481 GluGlnAlaArgGlnLeuLysAsnAsnProPheLysPheValProProSerGln 500
QY      1501 CAGTCTCCGAGGGCTGTGGCT 1521
Db      501 GlnSerProArgAlaValAla 507

RESULT 11
ABG71266
ID      ABG71266 standard; protein; 481 AA.
AC      ABG71266;
DT      17-DEC-2002 (first entry)
XX      Glycine max (Soybean) var. Dare protein.
DE      Soybean; Glycinin; atomic coordinate data; processability; soya protein;
KW      Dare; protein co-ordinate data.
XX      Glycine max.
OS      Glycine max.
XX      JP2002193996-A.
XX      10-JUL-2002.
XX      21-DEC-2000; 2000JP-00405097.
XX      21-DEC-2000; 2000JP-00405097.
XX      (KYOU ) UNIV KYOTO.
XX      WPI; 2002-685438/74.
XX      N-PSDB; ABS55193.
XX      Glycinin, beta-conglycinin and proglycinin, their crystal structures,
PT      three dimensional coordinates, three dimensional structured and models
PT      and their uses.
XX      Disclosure; Page 1273-1274; 1298pp; Japanese.
XX      The present invention relates to a new Glycinin characterised by the
CC      atomic coordinate data fully defined in the specification. The structure
CC      can be used for improving processability of soya protein. The present
CC      amino acid sequence represents the Glycine max (Soybean) var. Dare
CC      protein, as described in the specification
XX      SQ      Sequence 481 AA;

Alignment Scores:
Pred. No.:      1 32e-128      Length:      481
Score:          1502.00      Matches:      297
Percent Similarity: 71.57%      Conservative: 68
Best Local Similarity: 58.24%      Mismatches: 91
Query Match:      55.82%      Indels:      54
DB:              5      Gaps:      7

US-10-728-051-3 (1-1524) x ABG71266 (1-481)

QY      1 CGGCAGCAACCGGAGGAGACGGGTCCAGTTCAGCGCTCAATGCGCAGAGACCTGAC 60
Db      23 ArgGluGlnProGlnAsnGluCysGlnIleGlnArgLeuAsnAlaLeuLysProAsp 42
QY      61 AATCGCATTAATCAGAGGGCGGTTACATTGAGACTTGGAAACCCCAACACGAGGATTC 120
Db      43 AsnArgIleGluSerGluGlyGlyPheIleGluThrTrpAsnProAsnAsnLysProPhe 62
QY      121 GAATGCGCGCGCTCGCCCTCTCTCGCTTAGTCTCTCGCGCGCAACGCCCTTCGTAGGCCT 180
Db      63 GlnCysAlaGlyValAlaLeuSerArgCysThrLeuAsnArgAsnAlaLeuArgPro 82

```

```

QY      181 TTCTACTCAATGCTCTCCCGAGGAGATCTTTCATCCAGCAAGGAGGATACATTTGGGTTG 240
Db      83 SerTyrThrAsnAlaProGlnGluIleTyrIleGlnGlnGlySerGlyIlePheGlyMet 102
QY      241 ATATTCCCTGCTGCTAGACACTATGAAGAGCCCTCACACAAAGTTCGTGCATCTCAG 300
Db      103 IlePheProGlyCysProSerThrProGluGluProGlnGlnLysGly-----Gln 119
QY      301 TCCAAAGACCAACCAAGACGTCTCCAAGGAGAAGACCAAAGCCACACGCAACGAGATAGT 360
Db      120 SerSerArgPro-----GlnAspArg 126
QY      361 CACCAGAGGTGCACCGTTTCGATGAGGTGATCTCATTCAGTTCCTCCACCGTGTGTGCT 420
Db      127 HisGlnLysIleTyrHisPheArgGluGlyAspLeuIleAlaValProThrGlyPheAla 146
QY      421 TTCTGGCTCTACACGACCAACGACACTGATGTGTGCTGCTGCTTCTTACTGACACCAAC 480
Db      147 TyrTrpMetTyrAsnAsnGluAspThrProValValAlaValSerLeuIleAspThrAsn 166
QY      481 AACACGCAACACGCTTCATCAGTTCCTCCAGGAGATTCAATTTGGCTGGGAACACGGAG 540
Db      167 SerPheGlnAsnGlnLeuAspGlnMetProArgArgPheTyrLeuAlaGlyAsnGlnGlu 186
QY      541 CAAGAGTTCCTTAAGGTACCAACGCAACAAAGCAGCAACAAAGCAGCAAGAAAGCTTACCAT 600
Db      187 GlnGluPheLeuGlnTyrGlnProGlnLysGlnGlnGly-----199
QY      601 AGCCCATACAGCCCGCAAGTTCAGCTAGACAAAGAGAGCGGTGNAATTTAGCCCTCGAGGA 660
Db      200 -----Gly 200
QY      661 CAGCACAGCCGACAGAACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 720
Db      201 ThrGlnSerGlnLysGlyLysArgGlnGlnGlnGluGluGluGluGluGlySerIleLeu 220
QY      721 AGCGGCTTCACCGCGGAGTTCCTGGAAACAAAGCTTCAGGTTGACGACGACGACGATAGTG 780
Db      221 SerGlyPheAlaProGluPheLeuGluHisAlaPheValVal---AspArgGlnIleVal 239
QY      781 CAACACCTTAAGAGCGCAGACCGAGAGTGAAGAGAGGAGGAGCCATTTGTGCATCTGAGGGA 840
Db      240 ArgLysLeuGlnGlyLysGlnGluGluGluGluGlyValIleValThrValLysGly 259
QY      841 GGCTCAGAAATCTTGAGCCCA-----GATAGAAAGACGACGTCGCCACGAAAGAGAG 891
Db      260 GlyLeuSerValIleSerProThrGluGluGlnGlnArgProGluGluGluGlu 279
QY      892 GAATACGATGAAGATGAATATGAATACGATGAAGAGGATAGAAGCGCTGCGAGGGAAGC 951
Db      280 LysProAsp-----CysAspGluLysAspLysHisCysGlnSerGlnSer 294
QY      952 AGAGGCGAGGGGAATGGTATTGAGAGACGATCTGCACCGCAAGTCTCTAAAGACATTT 1011
Db      295 Arg-----AsnGlyIleAspGluThrIleCysThrMetArgLeuArgHisAsnIle 311
QY      1012 GGTAGAAACAGATCCCTCAGCATCTACAACCTCTCAAGCTGGTTCACCTAAAACGCGCAAC 1071
Db      312 GlyGlnThrSerSerProAspIlePheAsnProGlnAlaGlySerIleThrAlaThr 331
QY      1072 GATCTCAACTTCTTAATACTTAGTGGCTTGGACCTAGTGTGAATATATGGAATCTCTAC 1131
Db      332 SerLeuAspPheProAlaLeuSerTrpLeuLysLeuSerAlaGlnPheGlySerLeuArg 351
QY      1132 AGGAATGCAATTTGCTCGCTCAGTACACCAACCAACGACGACGACGACGACGACGACGAC 1191
Db      352 LysAsnAlaMetPheValProHisTyrAsnLeuAsnAlaAsnSerIleIleTyrAlaLeu 371
QY      1192 AGGGACGCGGCTCACCTGCAAGTCTGTGGACAGCAACGCGCAACGAGAGTGTACGACGAGGAG 1251
Db      372 AsnGlyArgAlaLeuValGlnValValAsnCysAsnGlyGluArgValPheAspGlyGlu 391

```


Db 43 AsnArgIleGluSerGluGlyGlyPheIleGluThrTrpAsnProAsnAsnLysProPhe 62
Qy 121 GAATGCCCGCGCTCGCCCTCTCTCGCTTAGTCTCCGCGCAACGCGCCCTTCGTAGGCCT 180
Db 63 GlnCysAlaGlyValAlaLeuSerArgCysThrLeuAsnArgAsnAlaLeuArgArgPro 82
Qy 181 TTCTACTCCATGCTCCCGAGGAGATCTTATCAGCAAGAGAGGGGATCTTTGGGTTG 240
Db 83 SerTyrThrAsnAlaProGlnGluIleTyrIleGlnGlnGlySerGlyIlePheGlyMet 102
Qy 241 ATATTCTCTGTTCTCTAGACACTATGAAGAGCTCACACACAAAGTTCGTGATCTCAG 300
Db 103 IlePheProGlyCysProSerThrPheGluGluProGlnGlnLysGly-----Gln 119
Qy 301 TCCCAAGACACCACAGACGCTCTCAAGGAGAGACCAAGCCACCAACAGACGATAGT 360
Db 120 SerSerArgPro-----GlnAspArg 126
Qy 361 CACAGAAGGTGACCGTTTCGATGAGGGTGATCTCATTTGGAGTTCCACCGGTGTGCT 420
Db 127 HisGlnLysIleTyrHisPheArgGluGlyAspLeuIleAlaValProThrGlyPheAla 146
Qy 421 TTCTGGCTCTACACGACACGACACGATGTTGCTGCTGCTTCTTACTTGACACCAAC 480
Db 147 TyrTrpMetTyrAsnAsnGluAspThrProValValAlaValSerLeuIleAspThrAsn 166
Qy 481 AACAAACACACACAGCTTGATCAGTCTCCAGGAGATTCAATTTGGCTGGGAACACCGAG 540
Db 167 SerPheGlnAsnGlnLeuAspGlnMetProArgArgPheTyrLeuAlaGlyAsnGlnGlu 186
Qy 541 CAGAGTTCTTAAGTACCAGCAACAAAGCAGACAAAGACGACGACGAGAGCTTACCATAT 600
Db 187 GlnGluPheLeuGlnTyrGlnProGlnLysGlnGly----- 199
Qy 601 AGCCCATACACCGCGCAAGTACGCTAGACAGAGCGTGAATTTAGCCCTCGAGGA 660
Db 200 -----Gly 200
Qy 661 CAGCACAGCGCAGACAGCAGCAGGACAAAGAAAGAAACGAGGTGGAAACATCTTC 720
Db 201 ThrGlnSerGlnLysGlyLysArgGlnGlnGluGlnGluGlnGlySerIleLeu 220
Qy 721 AGCGGCTTACGCGGAGTTCCTGGAAACAAGCTTCCAGGTTGACGACGACGATAGTG 780
Db 221 SerGlyPheAlaProGluPheLeuGluHieAlaPheValVal---AspArgGlnIleVal 239
Qy 781 CAAACCTAGACGCGGACGACGAGTCAAGAGAGGAGCCATTTGACAGTTCAGGGA 840
Db 240 ArgLysLeuGlnGlnGluAsnGluGluGluGluGlyAlaIleValThrValLysGly 259
Qy 841 GGCCTCAGAATCTTGAGCCCA-----GATAGAAAGAGCGTCCGCGACGAAGAGAG 891
Db 260 GlyLeuSerValIleSerProThrGluGluGlnGlnArgProGluGluGluGlu 279
Qy 892 GAATACGATGAAGATGAATATGAATACGATGAAGAGATGAAGCGGTGGCAGGGGAAGC 951
Db 280 LysProAsp-----CysAspGluLysAspLysHisCysGlnSerGlnSer 294
Qy 952 AGAGCGAGGGGAATGTTATTAAGACAGCATCTGCACCGCAAGTGTCTAAAGAAACATT 1011
Db 295 Arg-----AsnGlyIleAspGluThrIleCysThrMetArgLeuArgHisAsnIle 311
Qy 1012 GGTAGAACAGATCCCTGACATCTACACCTCAAGCTGTTCACTCAAACTGCCAAC 1071
Db 312 GlyGlnThrSerSerProAspIlePheAsnProGlnAlaGlySerIleThrAlaThr 331
Qy 1072 GATCTCAACCTTCTAATCTAGTGGCTTGGACCTAGTGTGTAATATGGAATCTCTAC 1131
Db 332 SerLeuAspPheProAlaLeuSerTrpLeuLysLeuSerAlaGlnPheGlySerLeuArg 351
Qy 1132 AGGAATCATGTTTGTGCTGCTACTACACACCAACGACGACATCATATATCGATTG 1191
Db 352 LysAlaMetPheValProHisTyrAsnLeuAsnAlaAsnSerIleIleTyrAlaLeu 371

Qy 1192 AGGGACGGGCTCACGTGCAAGTCTGGAGCAACGCGCAACAGAGTGTACGACGAGG 1251
Db 372 AsnGlyArgAlaLeuValGlnValValAsnCysAsnGlyGluArgValPheAspGlyGlu 391
Qy 1252 CTTCAAGAGGTCACGTGCTTGGTGGTCCACACAACTTCGCGCTCGTGGAAAGTCCAG 1311
Db 392 LeuGlnGluGlyGlnValLeuIleValProGlnAsnPheAlaValAlaAlaArgSerGln 411
Qy 1312 AGCGAGAACTTCGAATACGTGGCATTAACAGACAGACTCAAGGCCAGCATAGCCAACTC 1371
Db 412 SerAspAsnPheGluTyrValSerPheLysThrAsnAspArgProSerIleGlyAsnLeu 431
Qy 1372 GCCGTGAAACTTCGTCATAGATAACCTCCGAGGAGGTGTTGCAAAATTCATATGGC 1431
Db 432 AlaGlyAlaAsnSerLeuLeuAsnAlaLeuProGluGluValIleGlnGlnThrPheAsn 451
Qy 1432 CTCAAGAGGAGCAGCAGGACGCTTACAGCAACCAACCCCTCAAGTTCCTCGTTCCA 1491
Db 452 LeuArgArgGlnGlnAlaArgGlnValLysAsnAsnAsnProPheSerPheLeuValPro 471
Qy 1492 CCCTCTCAGCAGTCTCCGAGGGCTGTGGCT 1521
Db 472 ProLysGluSerGlnArgArgValAla 481
RESULT 14
ADG43988
ID ADG43988 standard; protein; 481 AA.
XX AC ADG43988;
XX XX
DT 26-FEB-2004 (first entry)
XX XX
DE G. max glycinin G3 subunit protein.
XX XX
KW oil content; plant; storage protein; seed-specific promoter; 2S-albumin;
KW 7S-globulin; 11S-globulin; 12S-globulin; zein-prolamine; transgenic;
KW oil production; fat production; free fatty acid production; food;
KW animal feed; pharmaceutical; fine chemical production; glycinin.
XX OS Glycine max.
XX PN WO2003077643-A2.
XX XX
PD 25-SEP-2003.
XX XX
PF 17-MAR-2003; 2003WO-EP002733.
XX XX
PR 20-MAR-2002; 2002DB-01012893.
XX XX
PA (BADI) BASF PLANT SCI GMBH.
XX PI Bauer J;
XX XX
DR WPI; 2004-011485/01.
DR N-PSDB; ADG43987.
XX PT Increasing total oil content of plants, useful e.g. as foods or animal
PT feeds, by reducing amount of storage proteins, particularly with double-
PT stranded interfering RNA.
XX XX
PS Claim 4; SEQ ID NO 28; 253pp; German.
XX XX
CC This invention describes a novel method for increasing the total oil
CC content of a plant by reducing the amount of at least one storage protein
CC in the plant (or its tissue, organs, parts or cells) and selecting plants
CC that have higher total oil content than starting plants. The storage
CC protein is suppressed by introducing antisense RNA, optionally combined
CC with a ribozyme, sense RNA that induces co-suppression, DNA-binding
CC factors directed against storage protein genes, viral sequences that
CC degrade storage protein RNA, constructs that induce homologous
CC recombination of endogenous storage protein genes or mutations into
CC storage protein genes. Most preferably a plant cell is stably transfected

CC with a recombinant expression construct, then regenerated to plants that express the incorporated sequence. The expression constructs particularly contain a seed-specific promoter and they are introduced into plants by standard methods, e.g. via Agrobacterium. The preferred storage proteins of the invention are 2S-albumens, 7S or 11S/12S-globulins or zein-prolamines. Transgenic organisms produced by the new method are used for production of oils, fats, free fatty acids or their derivatives, useful as foods, animal feeds, pharmaceuticals and fine chemicals. This sequence CC represents a storage protein used to illustrate the method of the invention.

XX
SQ Sequence 481 AA;

Alignment Scores:

Pred. No.: 1,32e-128 Length: 481
Score: 1502.00 Matches: 297
Percent Similarity: 71.57% Conservative: 68
Best Local Similarity: 58.24% Mismatches: 91
Query Match: 55.82% Indels: 54
DB: 8 Gaps: 7

US-10-728-051-3 (1-1524) x ADG43988 (1-481)

```

QY 1 CGCAGACACCGGAGGAGCGGTCCAGTTCAGCGCTCAATGGCAGAGACCTGAC 60
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23 ArgGluGlnProGlnGlnAsnGluCysGlnIleGlnArgLeuAsnAlaLeuLysProAsp 42
QY 61 AATCGCATTAATCAGAGGCGGTTCATTCAGACTTGGAACTCCCAACCAACAGGAGTTC 120
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
43 AsnArgIleGlnSerGluGlyGlyPheIleGluThrTrpAsnProAsnLysProPhe 62
QY 121 GAATGCGCGCGTCTCTCGCTTAGCTCTCGCTAGCTCTCGCGCGCAACGCGCTTCGTAGCCT 180
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
63 GlnCysAlaGlyValAlaLeuSerArgCysThrLeuAsnArgAsnAlaLeuArgPro 82
QY 181 TTCTACTCCAACTGCTCCAGGAGACTTTCATCCAGCAGAGGAGGAGTACTTTGGTTC 240
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
83 SerTyrThrAsnAlaProGlnGluIleTyrIleGlnGlnGlySerGlyIlePheGlyMet 102
QY 241 ATATTCCTCGTGTCTCTGACACTATGAGAGCGCTCACACAAAGCTCTCGATCTCAG 300
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103 IlePheProGlyCysProSerThrPheGluGluProGlnGlnLysGly-----Gln 119
QY 301 TCCCAAGACACCAAGACTCTCCAGGAGAGACCAAGCCACAGCAACAGAGATAGT 360
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
120 SerSerArgPro-----GlnAspArg 126
QY 361 CACCAGAGGTGCACCGTTTCGATAGGCTGATCTCATTCAGTCTCCACCGGTGTGCT 420
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
127 HisGlnLysIleTyrHisPheArgGluGlyAspLeuIleAlaValProThrGlyPheAla 146
QY 421 TTCTGGCTCTACACGACCAACGACACTGATGTGTGTGCTGTTCTCTTACTGACACCAAC 480
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
147 TyrTrpMetTyrAsnAsnGluAepThrProValValAlaValSerLeuIleAspThrAsn 166
QY 481 AACACGACACCAACGAGTGTAGTATGTTCCCGCAGGAGATTCATTTGGTGGGAACAGGAG 540
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 SerPheGlnAsnGlnLeuAspGlnMetProArgArgPheTyrLeuAlaGlyAsnGlnGlu 186
QY 541 CAAGAGTCTTAAAGTTACAGCAACAAAGCAGACAGACAGAGCAAGAGCTTACCATAT 600
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
187 GlnGluPheLeuGlnTyrGlnProGlnLysGlnGlnGly-----Gly 199
QY 601 AGCCCATACAGCCGCAAGTCCAGCTTAGACAAGAGCGGTGAATTTAGCCCTCGAGGA 660
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
200 -----Gly 200
QY 661 CAGCAGACGCGCAGAGAACGAGCAGGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201 ThrGlnSerGlnLysGlyLysArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 220
QY 721 AGCGGCTTCACGCGGAGTTCCTGGAAACAGACCTTCCAGGTTTCAGACAGACAGATAGTG 780
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

DB 221 SerGlyPheAlaProGluPheLeuGluHisAlaPheValVal---AspArgGlnIleVal 239
QY 781 CAAACCTTAAGAGCGGAGAGCCGAGAGTGAAGAGAGGAGGCCATTGTGACAGTGGGGA 840
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
240 ArgLysLeuGlnGlnGlyLeuAsnGluGluGluGlyValAlaIleValThrValLysGly 259
QY 841 GCGCTCAGAACTTGTAGGCCCA-----GATAGAAAGAGAGCTGCCACGAGAGAG 891
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
260 GlyLeuSerValIleSerProThrGluGluGlnGlnArgProGluGluGluGlu 279
QY 892 GAATACGATGAAGTGAATATGATGATGAGAGATGAAGAGCGGTGGCAGGGAAGC 951
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
280 LysProAsp-----CysAspGluLysAspLysHisCysGlnSerGlnSer 294
QY 952 AGAGCAGCGGGAATGTTATTGAAGAGACGATCTGCACCGCAGTCTCTAAAGAACATT 1011
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
295 Arg-----AsnGlyIleAspGluThrIleCysThrMetArgLeuArgHisAsnIle 311
QY 1012 GGTAGAAACAGATCCCTGACATCTACACCTCAAGCTGGTTCACCTCAAAACTGCCAAC 1071
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
312 GlyGlnThrSerSerProAspIlePheAsnProGlnAlaGlySerIleThrThrAlaThr 331
QY 1072 GATCTCAACCTTCTAATACTTAGTGGCTTGACCTAGTGTGCTGAATATGGAATCTCTAC 1131
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
332 SerLeuAspPheProAlaLeuSerTrpLeuLysLeuSerAlaGlnPheGlySerLeuArg 351
QY 1132 AGGAATGCACTTGTGTGCTCACTACACCAACGACGACGACGATCATATATCATGATTG 1191
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
352 LysAsnAlaMetPheValProHisTyrAsnLeuAsnAlaAsnSerIleIleTyrAlaLeu 371
QY 1192 AGGGAGCGGCTCAGTGCAGTCTGGACACGACGACGACGACGACGATGTACGACGAGGAG 1251
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
372 AsnGlyArgAlaLeuValGlnValValAsnCysAsnGlyGluArgValPheAspGlyGlu 391
QY 1252 CTTCAAGAGGTCACGTGCTGTGTGGTCCACAGAACTTCGCGCTCGCTGGAAGTCCCAAG 1311
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
392 LeuGlnGluGlyGlnValLeuIleValProGlnAsnPheAlaValAlaAlaArgSerGln 411
QY 1312 AGCGAAGACTTCGAATACGTGGCATTCAGACAGACTCAAGGCCAGCATAGCAACCTC 1371
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
412 SerAspAsnPheGluTyrValSerPheLysThrAsnAspArgProSerIleGlyAsnLeu 431
QY 1372 GCGGTGAAACTCCGTACATAGTAACTCGCGGAGGAGGTGTTGCCAAATTCATATGCG 1431
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
432 AlaGlyAlaAsnSerLeuLeuAsnAlaLeuProGluGluValIleGlnGlnThrPheAsn 451
QY 1432 CTCCAAGGAGCAGGCAAGCGAGCTTAAGAACACCAACCCCTTCAAGTCTTCGTTCCA 1491
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
452 LeuArgArgGlnGlnAlaArgGlnValLysAsnAsnAsnProPheSerPheLeuValPro 471
QY 1492 CCGTCTCAGCAGTCTCCGAGGCGTGTGGCT 1521
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
472 ProLysGluSerGlnArgArgValValAla 481

```

RESULT 15

ABG71265
ID ABG71265 standard; protein; 485 AA.

XX
AC ABG71265;

XX
DT 17-DEC-2002 (first entry)

XX
DE Glycine max (Soybean) var. Shiotsurunoko protein #2.

XX
KW Soybean; Glycinin; atomic coordinate data; processability; soya protein;
Shiotsurunoko; protein co-ordinate data.

XX
OS Glycine max.

XX
PN JP2002193996-A.

XX
PD 10-JUL-2002.

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 8, 2005, 03:45:11 ; Search time 257.329 Seconds
(without alignments)
6065.459 Million cell updates/sec

Title: US-10-728-051-3
Perfect score: 2691
Sequence: 1 CGGCGACGAGGAGGAGAA.....ctccgaggctgtgccttaa 1524

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 3224756

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlbp
-Q=/cgn2_1/USPTO.spool_p/US10728051/runat 07042005 125045 14049/app query.fasta_1.4757
-DB=Uniprot -QFMT=faстан -SUFFIX=rup -MINMATCH=0.1 -LOOPCT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pept -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcio -NORM=ext -HEA=SIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10728051 @CGN 1.1 874 @runat 07042005 125045 14049 -NCPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DLEXT=7

Database : Uniprot_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2673	99.3	507	2	082580
2	2479	92.1	529	2	Q9F211
3	2455	91.2	536	2	Q647H4
4	2451.5	91.1	537	2	Q647H3
5	2444	90.8	536	2	Q8LKN1
6	2436	90.5	536	2	Q6T2T4
7	2421	90.0	530	2	Q9SQH7
8	1834	68.2	510	2	Q61WG5
9	1651	61.4	484	2	Q647H2
10	1502	55.8	481	1	GLC3_SOYBN
11	1499.5	55.7	485	1	GLC2_SOYBN
12	1496	55.6	495	1	GLC1_SOYBN
13	1489	55.3	481	2	Q852U5
14	1488.5	55.3	482	2	Q852U4
15	1457.5	54.2	498	2	Q41702
16	1455	54.1	520	1	LEG2_PEA

17	1443	53.6	517	2	Q9T0P5
18	1440	53.5	517	1	LEGA_PEA
19	1431.5	53.2	482	2	Q41676
20	1414.5	52.6	500	2	Q99304
21	1375	51.1	497	2	Q03971
22	1186	44.1	496	2	Q9SMJ4
23	1110.5	41.3	515	2	Q8W1C2
24	1074.5	39.9	488	2	Q41128
25	1069.5	39.7	551	2	Q43607
26	1059	39.4	542	2	Q8LK20
27	1049	39.0	560	2	Q9S9D0
28	1034.5	38.4	457	2	Q8G2P6
29	1032	38.4	503	1	LEGI_PEA
30	999.5	37.1	566	2	Q24294
31	998	37.1	485	2	Q41703
32	996.5	37.0	484	1	LEGA_VICFA
33	986	36.6	517	2	Q39922
34	986	36.6	517	2	Q7GC77
35	985	36.6	517	2	Q9SB12
36	981.5	36.5	564	2	Q43673
37	980	36.4	517	2	P93707
38	980	36.4	517	2	P93708
39	978.5	36.4	219	2	Q8LL03
40	975	36.2	563	2	Q9SB11
41	974	36.2	563	2	Q39921
42	973.5	36.2	562	1	GLC4_SOYBN
43	968	36.0	504	2	Q43608
44	951.5	35.4	662	2	Q647H1
45	946.5	35.2	562	2	Q43452

ALIGNMENTS

RESULT 1

082580	PRELIMINARY;	PRT;	507 AA.
ID	082580	PRELIMINARY;	PRT;
AC	082580;		
DT	01-NOV-1998 (TEMBLrel. 08, Created)		
DT	01-NOV-1998 (TEMBLrel. 08, Last sequence update)		
DT	01-MAR-2004 (TEMBLrel. 26, Last annotation update)		
DE	Glycinin (Fragment).		
GN	Name=Arah3;		
OS	Arachis hypogaea (Peanut).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;		
OC	eurosid1; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;		
OC	Arachis.		
OX	NCBI_TaxID=3818;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=99146968; PubMed=10021462;		
RA	Rabjohn P., Helm E.M., Stanley J.S., West C.M., Sampson H.A.,		
RA	Burks A.W., Banton G.A.;		
RT	"Molecular cloning and epitope analysis of the peanut allergen Ara h		
RT	3.";		
RL	J. Clin. Invest. 103:535-542(1999).		
DR	EMBL; AF093541; AAC63045.1; -;		
DR	HSSP; P04776; IUD1.		
DR	GO; GO:0045735; F:nutrient reservoir activity; IEA.		
DR	InterPro; IPR006045; Cupin.		
DR	InterPro; IPR007113; Cupin-region.		
DR	InterPro; IPR011051; RmlC_like_cupin.		
DR	InterPro; IPR006044; Seedstore_11s.		
DR	Pfam; PF001190; Cupin; 2.		
DR	PRINTS; PR00439; 11SGLOBULIN.		
DR	PROSITE; PS00305; 11S_SEED_STORAGE; UNKNOWN_1.		
FT	NON_TER		
SQ	SEQUENCE 507 AA; 58349 MW; P3FB38BC3CB82DED CRC64;		

Alignment Scores:

Pred. No.: 5,12e-179 Length: 507
Score: 2673.00 Matches: 507
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00%		Mismatches: 0	
Query Match: 99.33%		Indels: 0	
DB: 2		Gaps: 0	
US-10-728-051-3 (1-1524) x 082580 (1-507)			
QY	1	CGGCAGCAACCGGAGAGAACGGTCCAGCTTCCAGCGCTCAATGCGCAGAGACCTGAC	60
DB	1	ArgGlnGlnProGluGluAsnAlaCysGlnPheGlnArgLeuAsnAlaGlnArgProASP	20
QY	61	AATCGCAATGAATCAGAGCGGCTTACATGAGACTTGGAAACCCCAACACAGAGATTC	120
DB	21	AsnArgileGluSerGluGlyGlyTyrIleGluThrTyrAsnProAsnAsnGlnGluPhe	40
QY	121	GAATGCGCGGCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	180
DB	41	GluCysAlaGlyValAlaLeuSerArgLeuValLeuArgArgAsnAlaLeuArgArgPro	60
QY	181	TTCTACTCTCAATGCTCCCGAGAGATCTTCATCCAGCAAGGAAGGGGATCTTTGGGTTG	240
DB	61	PheTyrSerAsnAlaProGlnGlnIlePheIleGlnGlnGlyArgGlyTyrPheGlyLeu	80
QY	241	ATATTCCTGGTGTCTAGACACTATGAAGAGCTTCACACAGAGTGTGATCTCAG	300
DB	81	IlePheProGlyCysProArgHisTyrGluGluProHisThrGlnGlyArgArgSerGln	100
QY	301	TCCCAAGACCAACAGAGCTCTCCAGGAGAGACCAAGCCCAACAGCAACGAGATAGT	360
DB	101	SerGlnArgProProArgArgLeuGlnGlyGluAspGlnSerGlnGlnGlnArgAspSer	120
QY	361	CACGAGAAGTGACCGCTTTCGATGAGGTGTCTCTATTGCAATGCCACCGGTGTGCT	420
DB	121	HisGlnIysValHisArgPheAspGluGlyAspLeuIleAlaValProThrGlyValAla	140
QY	421	TTCTGGCTCTACACGACACGACACTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTG	480
DB	141	PheTyrLeuTyrAsnAspHisAspThrAspValValAlaValSerLeuThrAspThrAsn	160
QY	481	AACAACGACACACAGCTTGATCATGTTCCCGAGAGATTCATTTGGCTGGGACACGGAG	540
DB	161	AsnAsnAspAsnGlnLeuAspGlnPheProArgArgPheAsnLeuAlaGlyAsnThrGlu	180
QY	541	CAAGAGTCTTTAAGTACCAGCAACAAAGCAGACAAAGCAGACGAAAGAGCTTACCAT	600
DB	181	GlnGluPheLeuArgTyrGlnGlnSerArgGlnSerArgArgSerLeuProTyr	200
QY	601	AGCCCATACAGCCGCAAGTACCTAGCTAGCAGACAGAGCGTCAATTTAGCCCTCAGGA	660
DB	201	SerProTyrSerProGlnSerGlnProArgGlnGluGluArgGluPheSerProArgGly	220
QY	661	CAGCAGCCGCGAGAGACGAGCAGACAGACAGACAGACAGACAGACAGACATCTTC	720
DB	221	GlnHisSerArgArgGluArgAlaGlyGlnGluGluGluGluGluGluGluGluGluPhe	240
QY	721	AGCGGTTCTACCGCGGAGTCTCTGGAACAAGCTTCCAGGTTCCAGCAGACAGATAGTG	780
DB	241	SerGlyPheThrProGluPheLeuGluGlnAlaPheGlnValAspArgGlnIleVal	260
QY	781	CAAAACCTAGAGCGGAGACCGAGAGTGTAGAGAGAGGAGCCATTCTGACAGTGAGGGA	840
DB	261	GlnAsnLeuArgGlyGluThrGluSerGluGluGluGluGluGluGluGluGluGluGly	280
QY	841	GGCTCAGATCTTTAGCCCGAGATAGAAAGAGACGTCGCCAGCAAGAGGAATACGAT	900
DB	281	GlyLeuArgileLeuSerProAspArgLysArgArgAlaAspGluGluGluTyrAsp	300
QY	901	GAGATGAATATGAATACGATGAGAGATAGAGCGTGGCAGGAGGAGCAGAGCAGG	960
DB	301	GluAspGluTyrGluTyrAspGluGluAspArgArgArgGlyArgGlySerArgGlyArg	320
QY	961	GGGATGATTTAGAGAGCGATCTGCACCGCAAGTCTCAAAAGACATTTGCTAGAAC	1020
DB	321	GlyAsnGlyIleGluGluThrIleCysThrAlaSerAlaLysLysAsnIleGlyArgAsn	340

QY	1021	AGATCCCTGACATCTACAAACCTCAAGCTGTTCTCACTCAAAACCTCAAGCTCAAC	1080
DB	341	ArgSerProAspIleTyrAsnProGlnAlaGlySerLeuIysThrAlaAsnAspLeuAsn	360
QY	1081	CTTCTAATCTAGTGTGGACCTAGTGTCTGAATATGAAATCTCTACAGGAATGCA	1140
DB	361	LeuLeuIleLeuArgTyrProSerAlaGluTyrGlyAsnLeuTyrArgAsnAla	380
QY	1141	TTGTTTGTGCTCACTACAAACCAACGACACACAGCTGATCATATATCGATTGAGGGG	1200
DB	381	LeuPheValAlaHisTyrAsnThrAsnAlaHisSerIleIleTyrArgLeuArgGlyArg	400
QY	1201	GCTCAGCTGCAAGTCTGACAGCAACCGCAACAGAGTGTACGACGAGGAGCTTCAAG	1260
DB	401	AlaHisValGlnValValAspSerAsnGlyAsnArgValTyrAspGluGluLeuGlnGlu	420
QY	1261	GCTCAGCTGCTGTGCTGCCACAGAACTTCGCGCTCGTGGAAAGTCCAGAGCAGAAC	1320
DB	421	GlyHisValLeuValValProGlnAsnPheAlaValAlaGlyLysSerGlnSerGluAsn	440
QY	1321	TTTCAATACGTGGCATTCAAGACACACTCAAGGCCAGCAGTACCAACCTCGCCGTGAA	1380
DB	441	PheGluTyrValAlaPheLysThrAspSerArgProSerIleAlaAsnLeuAlaGlyGlu	460
QY	1381	AACTCCGTCATAGATAAATCTGCCGAGGAGGTGTTGCAAAATTCATATGCGCTCCAAAG	1440
DB	461	AsnSerValIleAspAsnLeuProGluGluValValAlaAsnSerTyrGlyLeuGlnArg	480
QY	1441	GAGCAGGACGAGCAGCTTAAGAACCAACCCCTCAAGTCTTCTGTTCCACCGCTCTAG	1500
DB	481	GluGlnAlaArgGlnLeuLysAsnAsnProPheLysPhePheValProProSerGln	500
QY	1501	CAGTCTCCGAGGCTGTGGCT	1521
DB	501	GlnSerProArgAlaValAla	507
RESULT 2			
ID	Q9FZ11	PRELIMINARY;	PRT; 529 AA.
AC	Q9FZ11		
DT	01-MAR-2001	(TRENBLrel. 16, Created)	
DT	01-MAR-2001	(TRENBLrel. 16, Last sequence update)	
DT	01-MAR-2004	(TRENBLrel. 26, Last annotation update)	
DE	Gly1.		
GN	Name=Gly1;		
OS	Arachis hypogaea (Peanut).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;		
OC	eurosid I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;		
OC	Arachis.		
OX	NCBI_TaxID=3818;		
RN	[1]_TaxID=3818;		
RP	SEQUENCE FROM N.A.		
RA	Jain A.K., Basha S.M.;		
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF125192; AAG01363.1; -		
DR	HSSP; P04776; 1UCX.		
DR	GO; GO:0045735; F.nutrient reservoir activity; IEA.		
DR	InterPro; IPR006045; Cupin.		
DR	InterPro; IPR007113; Cupin_region.		
DR	InterPro; IPR011051; RmC_like_cupin.		
DR	Pfam; PF00190; Cupin; 2.		
DR	PRINTS; PR00439; 11SGLOBULIN.		
DR	PROSITE; PS00305; 11S SEED STORAGE; UNKNOWN 1.		
SQ	SEQUENCE 529 AA; 60448 MW; BB6F25BC1D6E06A1 CRC64;		
Alignment Scores:			
Pred. No.:	2,11e-165	Length:	529
Score:	2479.00	Matches:	475
Percent Similarity:	95.28%	Conservative:	9
Best Local Similarity:	93.50%	Mismatches:	20

Query Match:	92.12%	Indels:	4
DB:	2	Gaps:	2
US-10-728-051-3 (1-1524) x Q9FZ11 (1-529)			
QY	1	CGCAGCAACCGGAGGAGACGGTGCAGTTCACGCGCTCAATGCGCAGACCTGAC	60
DB	25	ArgGlnGlnProGluGluAenAlaCysGlnPheGlnArgLeuAenAlaGlnArgProAsp	44
QY	61	AATCGCATTTGAATCAGAGGCGGTTCATTGAGACTTTGGAAACCCCAACACACAGGAGTTC	120
DB	45	AsnArgLeuGluSerGluGlyGlyTyrIleGluThrTrpAsnProAenGlnGluPhe	64
QY	121	GAATGCGCGGCTGCGCTCTCTCGCTTAGTCTCTCGCGCGAAGCGCCCTTCGTAGCCT	180
DB	65	GluCysAlaGlyValAlaLeuSerArgLeuValLeuArgAsnAlaLeuArgArgPro	84
QY	181	TTCTACTCCAAATGCTCCCAAGGAGATCTTCATCCAGCAAGGAAGGATATTTGGGTG	240
DB	85	PheTyrSerAsnAlaProGlnGluIlePheIleGlnGlyArgGlyTyrPheGlyLeu	104
QY	241	ATATTCCCTGCTGTCTAGACACTATGAAGAGCTCACACAAAGTCTCGATCTCAG	300
DB	105	IlePheProGlyCysProSerThrTyrGluGluProAlaGlnGlnArgGlnGln	124
QY	301	TCCCAAGACCAACAGAGCTCTCCAAAGGAGAGACCAAAAGC---CAACAGCAACGAGAT	357
DB	125	SerGlnArgAlaProArgArgPheGluGlyGluAspGlnSerGlnGlnGlnGlnAsp	144
QY	358	AGTCACCAAGAGTGACCGCTTCGATGAGGTGATCTCATTCAGTCTCCACCGGTGT	417
DB	145	SerHisGlnLysValArgArgPheAspGluGlyAspLeuIleAlaValProThrGlyVal	164
QY	418	GCTTTCTGGCTCTACACAGACCAACGACACTGATGTTGCTGTTCTTCTTACTGACAC	477
DB	165	AlaLeuTrpMetPheAsnAspHisAspThrAspValValAlaValSerLeuThrAspThr	184
QY	478	AACAACCAACCAACAGCTTGATCAGTTCCTCCAGGAGATTCATTTGGCTGGGAACAG	537
DB	185	AsnAsnAsnAspAsnGlnLeuAspGlnPheProArgArgPheAsnLeuAlaGlyAsnHis	204
QY	538	GACCAAGAGTCTTAAAGTACCAGCAACAAAGCAGACAAAGCAGACGAAAGCTTACCA	597
DB	205	GluGlnGluPheLeuArgTyrGlnGln-----GlnSerArgArgSerLeuPro	221
QY	598	TATAGCCCATACAGCCGCAAGTACGCTAGACAGAGAGCTGATTTAGCCCTCGA	657
DB	222	TyrSerProTyrSerProGlnSerGlnProArgGlnGluArgGluPheSerProArg	241
QY	658	GGACAGCACCGCCAGAACCGAGCAGGACAAAGAAAGAAACGAAGTGGAAACATC	717
DB	242	GlyGlnHisSerArgArgGluArgAlaGlyGlnGluGluAsnGluGlyGlyAsnIle	261
QY	718	TTACAGCGCTTACCGCGGAGTCTCTGGAACAGCTTCAGGTGACGACAGACAGATA	777
DB	262	PheSerGlyPheThrProGluPheLeuAlaGlnAlaPheGlnValAspAspArgGlnIle	281
QY	778	GTGCAAAACCTTAAGCGCAGACCGAGAGTGAAGAGGAGGAGCATTTGTGACAGTGGG	837
DB	282	ValGlnAsnLeuArgGlyGluAsnGluSerGluGluGlyAlaIleValThrValLys	301
QY	838	GGAGGCTTCAGATCTTTGAGCCCAAGATAGAAAGAGAGCTGCCGCAAGAGGAGATAC	897
DB	302	GlyGlyLeuArgIleLeuSerProAspArgLysArgGlyAlaAspGluGluGluGluTyr	321
QY	898	GATCAAGATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	957
DB	322	AspGluAspGluTyrGluTyrAspGluGluAspArgArgArgGlyArgGlySerArgGly	341
QY	958	AGGGGGAATGTTTGAAGAGACCATCTGCACCGCAAGTGTCTAAAGAAACATTGGTGA	1017
DB	342	ArgGlyAsnGlyIleGluGluThrIleCysThrAlaSerValLysLysAsnIleGlyArg	361
QY	1018	AACAGATCCCTCAGCATCTACACCTCAAGCTGGTTCCTCACTCAAACTGCCACGATCTC	1077
DB	362	AsnArgSerProAspIleTyrAsnProGlnAlaGlySerLeuLysThrAlaAsnAspLeu	381
QY	1078	AACCTTCTAATACTAGGTGGCTTGACCTAGTGTGCTGAATATATGAAATCTCTACAGAAAT	1137
DB	382	AsnLeuLeuIleLeuArgTTrpLeuGlyLeuSerAlaGluTyrGlyAsnLeuTyrArgAsn	401
QY	1138	GCATTCTTTGCTCGCTCCTACCAACCAACGACACAGCATCATATATCGATTGAGGGA	1197
DB	402	AlaLeuPheValProHisTyrAsnThrAsnAlaHisSerIleIleTyrAlaLeuArgGly	421
QY	1198	CGGGCTCAGTCCCAAGTCTGCGACCAACGCAACAGAGTGTACGACGAGGACTTCAA	1257
DB	422	ArgAlaHisValGlnValValAspSerAsnGlyAsnArgValTyrAspGluGluLeuGln	441
QY	1258	GAGGGTCAGTGTGCTGCTGCGTGCACAGAACTTCGCGCTCGCTGAAAGTCCACGAGCGAG	1317
DB	442	GluGlyHisValLeuValValProGlnAsnPheAlaValAlaGlyLysSerGlnSerAsp	461
QY	1318	AATCTCGAATACGTGCTTCAAGCAGACTCAAGCCCGCAGCATGCCAACCTTCATATGCGCTCAA	1377
DB	462	AsnPheGluTyrValAlaPheLysThrAspSerArgProAsnIleAlaAsnPheAlaGly	481
QY	1378	GAATACTCGTCATAGATAACCTGCCGAGGAGGTGTTGCCAATTCATATGCGCTCAA	1437
DB	482	GluAsnSerIleIleAspAsnLeuProGluGluValValAlaAsnSerTyrGlyLeuPro	501
QY	1438	AGGGACGAGCAGGAGGAGCTTAAAGAACAAACACCCCTCAAGTCTTCGTTCCACCGTCT	1497
DB	502	ArgGluGlnAlaArgGlnLeuLysAsnAsnAsnProPheLysPheValProProSer	521
QY	1498	CAGCAGTCTCCGAGGCTGTGCT 1521	
DB	522	GlnGlnSerLeuArgAlaValAla 529	
RESULT 3			
ID	Q647H4	PRELIMINARY;	PRT; 536 AA.
AC	Q647H4;		
DT	25-OCT-2004	(TrEMBLrel. 28, Created)	
DT	25-OCT-2004	(TrEMBLrel. 28, Last sequence update)	
DT	25-OCT-2004	(TrEMBLrel. 28, Last annotation update)	
DE	Arachin Ahy-1.		
OS	Arachis hypogaea (peanut).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;		
OC	eurosid I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;		
OC	Arachis.		
OX	NCBI_TaxID=3818;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Yan Y., Wang L., Huang S.;		
RT	"cDNA clone of peanut seed storage protein gene."		
RL	Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF722685; AAU21490.1; -.		
SQ	SEQUENCE 536 AA; 61505 MW; 30DBF638719AEB78 CRC64;		
Alignment Scores:			
Pred. No.:	1,02e-163	Length:	536
Score:	2455.00	Matches:	471
Percent Similarity:	93.99%	Conservative:	14
Best Local Similarity:	91.28%	Mismatches:	19
Query Match:	91.23%	Indels:	12
DB:	2	Gaps:	3
US-10-728-051-3 (1-1524) x Q647H4 (1-536)			
QY	1	CGCAGCAACCGGAGGAGACGGTGCAGTTCACGCGCTCAATGCGCAGACCTGAC	60
DB	24	ArgGlnGlnProGluGluAenAlaCysGlnPheGlnArgLeuAenAlaGlnArgProAsp	43
QY	61	AATCGCATTTGAATCAGAGGCGGTTCATTGAGACTTTGGAAACCCCAACACAGGAGTTC	120

```
Db 44 AsnArgIleGluSerGluGlyGlyTyrIleGluThrTrpAsnProAsnAsnGlnGluPhe 63
QY 121 GAATGGCGGGCTCGCTCTCTCGCTTAGTCTCCGCGGAAACCCCTTCGTAGGCT 180
Db 64 GluCyAlaGlyValAlaLeuSerArgLeuValLeuArgArgAsnAlaLeuArgPro 83
QY 181 TTCTACTCCAATGCTCCAGAGAGATCTTCATCCAGCAAGGAGGATCTTTGGTTG 240
Db 84 PheTyrSerAsnAlaProGlnGluIlePheIleGlnGlnGlyArgGlyTyrPheGlyLeu 103
QY 241 ATATTCCTCGTGTCTAGACACTATGAAGAGCTTCACACAAAGTTCGTGATCTCAG 300
Db 104 IlePheProGlyCysProSerThrTyrGluGluProAlaGlnGlnGlyArgGlnHisGln 123
QY 301 TCCCAAGACCAACAGAGCTCCCAAGAGAGAGCAACCAAGCAACAGCAAGATAGT 360
Db 124 SerGluArgProArgArgPheGlnGlyGlnAspGlnSerGlnGlnGlnGlnAspSer 143
QY 361 CACCAGAAGTGCACCGTTTCGATGAGGTGATCTCATTCAGATTCACACCGGTGTTCT 420
Db 144 HisGlnLysValHisArgPheAspGluGlyAspLeuIleAlaValProThrGlyValAla 163
QY 421 TTCTGCTCTACAAACAGACAGACTGATGTTGTTGTTCTCTTACTGACACCAAC 480
Db 164 PheTrpMetTyrAsnAspHisAspThrAspValValAlaValSerLeuThrAspThrAsn 183
QY 481 AACCAACGACACAGCTTGATCAGTTCCCGAGAGATTCATTTGCTGGCAACAGGAG 540
Db 184 AsnAsnAspAsnGlnLeuAspGlnPheProArgArgPheAsnLeuAlaGlyAsnHisGlu 203
QY 541 CAAGAGTCTTTAAGTACCAGCAACAAAGCAGCAAGCAAGAGAGAGAGAGAGAGAG 600
Db 204 GlnGluPheLeuArgTyrGlnGln-----GlnSerArgArgSerLeuProTyr 220
QY 601 AGCCCATACGCGCGAACTGAGCTAGCTAGACAGAGAGAGAGAGAGAGAGAGAGAG 660
Db 221 SerProTyrSerProGlnThrGlnProLysGlnGluAspArgGluPheSerProArgGly 240
QY 661 CAGCAGACCGCGAGAGCAAGCAGCAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
Db 241 GlnHisGlyArgArgGluArgAlaGlyGlnGlnGluAsnGluGlyGlyAsnIlePhe 260
QY 721 AGCGGCTTACGCGGAGTTCCTGGAAACAAGCTTCAGCTTCAGCAGCAGCAGATAGTG 780
Db 261 SerGlyPheThrProGluPheLeuAlaGlnAlaPheGlnValAspArgGlnIleLeu 280
QY 781 CAAACCTAAGAGCGAGACCGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
Db 281 GlnAsnLeuArgGlyGluAsnGluSerAspGluGlnGlyAlaIleValThrValArgGly 300
QY 841 GGCCTCAGAATCTTGAGCCAGATAGAAAGAGA-----CGTGGCGAGCA 885
Db 301 GlyLeuArgIleLeuSerProAspArgLysArgGlnGlnTyrGluArgProAspGlu 320
QY 886 GAAGAGAAATCAGATGAAGATGAATATGAATATGATGATGATGATGATGATGATGAT 933
Db 321 GluGluGluTyrAspGluAspGluTyrGluTyrAspGluGluGluGluGluGluGlu 340
QY 934 AGCGTGGCAGGGAGCAGCAGCAGCGGGAAGTATTGAAGAGAGATCTGACCGCA 993
Db 341 ArgArgGlyArgGlySerArgGlySerGlyAsnGlyIleGluGluThrIleCysThrAla 360
QY 994 AGTGCTAAAAAGAACATTGGTAGAAAACAGATCCCTTCACATCTACAAACCTCAAGCTGGT 1053
Db 361 SerPheLysLysAsnIleGlyArgAsnArgSerProAspIleTyrAsnProGlnAlaGly 380
QY 1054 TCACTCAAACTGCAACAGATCTCAACCTTCTAATACTTAGTGGCTTGACACCTAGTCT 1113
Db 381 SerLeuLysThrAlaAsnGluLeuAsnLeuIleLeuArgTrpLeuGlyLeuSerAla 400
QY 1114 GAATATGAAATCTCTACAGAAATGCATTGTTGTCCTCACTACACACACACAGCAC 1173
```

```
Db 401 GluTyrGlyAsnLeuTyrArgAsnAlaLeuPheValProHisTyrAsnThrAsnAlaHis 420
QY 1174 AGCATCATATATCGATTGAGGGGACGGCTCACGTGCAAGTCGTGACAGCAACGGCAAC 1233
Db 421 SerIleIleTyrAlaLeuArgIleValGlnValValAspSerAsnGlyAsp 440
QY 1234 AGAGTGTACGACGAGGAGCTTCAAGAGGTCACGTGCTTGTGTCGACAGACTTCGCC 1293
Db 441 ArgValPheAspGluGluGlnGluGlyHisValLeuValValProGlnAsnPheAla 460
QY 1294 GTTCGTGCAAAAGTCCAGAGCGAGAACTTCAATACGTGGCATTCAGACAGACTCAAGG 1353
Db 461 ValAlaGlyLysSerGlnSerGluAsnPheGluTyrValAlaPheLysThrAspSerArg 480
QY 1354 CCCAGCATAGCAACTCCCGCGTGAAACTCCGTCATAGATAAACCTCCGAGAGAGGTG 1413
Db 481 ProSerIleAlaAsnLeuAlaGlyGluAsnSerPheIleAspAsnLeuProGluGluVal 500
QY 1414 GTTGAATATTCATATGGCTCCAAAGGAGCAGCAGCAGCTTAAGCAACACACACCCC 1473
Db 501 ValAlaAsnSerTyrGlyLeuProArgGluGlnAlaArgGlnLeuLysAsnAsnPro 520
QY 1474 TTCAAAGTTCTCGTTCACCGCTCTCAGCAGTCTCCGAGGCTGTGGCT 1521
Db 521 PheLysPhePheValProSerGluGlnSerLeuArgAlaValAla 536
RESULT 4
Q647H3 PRELIMINARY; PRT; 537 AA.
AC Q647H3;
DT 25-OCT-2004 (TReMBLrel. 28, Created)
DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)
DE Arachin Ahy-2.
OS Arachis hypogaea (Peanut).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eusoids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;
OC Arachis.
OX NCBI_TaxID=3818;
RN [1]
RP SEQUENCE FROM N.A.
RA Yan Y., Wang L., Huang S.;
RT "cDNA clone of peanut seed storage protein gene.";
RL Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AY722686; AAU21491.1; -.
SQ SEQUENCE 537 AA; 61531 MW; 3BDD034DEA159657 CRC64;
```

Alignment Scores:
Pred. No.: 1,8e-163 Length: 537
Score: 2451.50 Matches: 474
Percent Similarity: 93.62% Conservative: 10
Best Local Similarity: 91.68% Mismatches: 20
Query Match: 91.10% Indels: 13
DB: 2 Gaps: 4

US-10-728-051-3 (1-1524) x Q647H3 (1-537)

```
QY 1 CGGCAGCAACCGAGAGAACCGTCCAGCTCCAGCGCTCAATCGCAGAGACTGAC 60
Db 24 ArgGlnGlnProGluGluAsnAlaCysGlnPheGlnArgLeuAsnAlaGlnArgProAsp 43
QY 61 ATTCGATTGAATCAGAGGCGGTTACATTGACTTGAACCCCAACACAGGAGTTC 120
Db 44 AsnArgLeuGluSerGluGlyGlyTyrIleGluThrTrpAsnProAsnAsnGlnGluPhe 63
QY 121 GAATGGCGGGCTCGCTCTCTCGCTTAGTCTCCGCGGCAACGGCTTCGTAGGCT 180
Db 64 GluCyAlaGlyValAlaLeuSerArgLeuValLeuArgArgAsnAlaLeuArgPro 83
QY 181 TTCTACTCCAATGCTCCCGAGAGATCTTCATCCAGCAAGGAGGATCTTTGGTTG 240
Db 84 PheTyrSerAsnAlaProGlnGluIlePheIleGlnGlnGlyArgGlyTyrPheGlyLeu 103
```

QY 241 ATATTCCCTGGTGTCTCTAGACACTATGAAGAGCCCTCACACAAGGTGCTCGATCTCAG 300
Db 104 IlePheProGlyCysProSerThrTyrgluGluProAlaGlnGlnGlyArgArgHisGln 123
QY 301 TCCCAAGACCCAAAGAGCTCTCCAGGAGAGAGCAAGC--CAACAGCAACAGAT 357
Db 124 SerGlnArgAlaProArgArgPheGluGlyGluSerGlnGlnGlnGlnAsp 143
QY 358 AGTCACCAAGGTGACCCGTTTCGATGAGGGTGATCTCATTCAGTCTCCACCGGTGTT 417
Db 144 SerHisGlnIysValArgArgPheAspGluGlyAspLeuIleAlaValProThrGlyVal 163
QY 418 GCTTCTGGCTCTACAGACACGACATGATGTTGTGCTGCTCTCTTACTGACACC 477
Db 164 AlaLeuTrpMetTyrAsnAspHisAspThrAspValValAlaValSerLeuThrAspThr 183
QY 478 AACAAACAGACACAGCTTGATCAGTTCCTCCAGGAGTTCATTTGGCTGGACACG 537
Db 184 AsnAsnAsnAspAsnGlnLeuAspGlnPheProArgPheAsnLeuAlaGlyAsnHis 203
QY 538 GAGCAAGAGTCTTTAAGGTACCAAGCAACAAAGCAGACAAAGCAGACAAAGCTTACCA 597
Db 204 GluGlnGluPheLeuArgTyrGlnGln-----GlnSerArgArgSerLeuPro 220
QY 598 TATAGCCCATACAGCCGCAAGTCAAGCTAGACAGAGAGCGTGAATTTAGCCCTCGA 657
Db 221 TyrSerProTyrSerProGlnSerGlnProArgGlnGluGluArgGluPheSerProArg 240
QY 658 GGNACACACAGCCGACAGACAGCAGCAGACAGACAGACAGACAGACAGACAGACATC 717
Db 241 GlyGlnHisSerArgArgGluArgAlaGlyGlnGlnGlnGlnGlnGlnGlnGln 260
QY 718 TTCAGCGGCTTCACGCGGAGTTCCTCGAACAAGCTTCAGGTTCCACACACAGACATA 777
Db 261 PheSerGlyPheThrProGluPheLeuAlaGlnAlaPheGlnValAspAspArgGlnIle 280
QY 778 GTGCAAAACCTAAGACGCGAGACCGAGAGTGAAGAAGAGAGGACCCATTGTGACAGTGG 837
Db 281 LeuGlnAsnLeuArgGlyGluAsnGluSerAspGluGlnGlyAlaIleValThrValArg 300
QY 838 GGNCGCTCAGATCTTGACCCAGATAGAACAGA-----CGTGGCCGAC 882
Db 301 GlyGlyLeuArgIleLeuSerProAspArgLysArgGlnGlnTyrGluArgProAsp 320
QY 883 GAAGAAGAGGAATACGATGAAGATGAATATGATACATGAAGAG-----GAT 930
Db 321 GluGluGluGluTyrAspGluAspGluTyrGluTyrAspGluGluArgGlnGlnAsp 340
QY 931 AGAAGCGTGGCAGGGGAAGCAGAGCGAGGGGAATGGTATTCAAGAGACGATCTGCACC 990
Db 341 ArgArgGlyArgGlySerArgGlyArgGlyAsnGlyIleGluGluThrIleCysThr 360
QY 991 GCAAGTGTCTAAAAGACATGTTGTAGAACAGATCCCTGCATCTCAACCCCTCAAGCT 1050
Db 361 AlaSerValIysIysAsnIleGlyArgAsnArgSerProAspIleTyrAsnProGlnAla 380
QY 1051 GGTTCCTCAAACTGCCAAGATCTCAACCTTCTAATCTTAGTGGCTTGGACCTAGT 1110
Db 381 GlySerLeuIysThrAlaAsnAspLeuAsnLeuIleLeuArgTrpLeuGlyLeuSer 400
QY 1111 GCTGAATATGAAATCTTACAGGAATGCAATGTTGTCCTCACTACCAACCAACGCA 1170
Db 401 AlaGluTyrGlyAsnLeuTyrArgAsnAlaLeuPheValProHisTyrAsnThrAsnAla 420
QY 1171 CACAGATCATATATGATTGAGGAGCGGGCTCAGTGCAGAACTGTGGACAGACGCG 1230
Db 421 HisSerIleIleTyrAlaLeuArgGlyArgAlaHisValGlnValValAspSerAsnGly 440
QY 1231 AACAGAGTGTACACGAGGAGCTTCAAGAGGTCAGTGGTCTGTCGTCACAGACTTC 1290
Db 441 AsnArgValTyrAspGluGluLeuGlnGlnGlyHisValLeuValProGlnAsnPhe 460

1291 GCGTCCGTGGAAAGTCCAGAGCGAGAACTTCGAATAGCTGCATTCAAGACAGACTCA 1350
Db 461 AlaValAlaGlyLysSerGlnSerAspAsnPheGluTyrValAlaPheLysThrAspSer 480
QY 1351 AGCCCCAGCATAGCCAACTCCGCGTGAAACTCCGTCATAGATAACTCCGCGGAGGAG 1410
Db 481 ArgProSerIleAlaAsnLeuAlaGlyGluAsnSerIleIleAsnLeuProGluGlu 500
QY 1411 GTGGTTGCAAAATTCATATGCGCTCCAAAGGAGCAGCAGGAGGAGGCTTAAAGAAACAAC 1470
Db 501 ValValAlaAsnSerTyrGlyLeuProArgGluGlnAlaArgGlnLeuLysAsnAsn 520
QY 1471 CCCTTCAAGTCTTCCGTTCCACCGTCTCAGCAGTCTCCGAGGCTGTGGCT 1521
Db 521 ProPheLysPheValProSerGlnGlnSerLeuGlyAlaValAla 537

RESULT 5
O8LKN1 PRELIMINARY; PRT; 538 AA.
ID Q8LKN1
AC Q8LKN1;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Allergen Arab3/Arab4.
OS Arachis hypogaea (peanut).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;
OC Arachis.
OX NCBI_TaxID=3818;
RN [1]
RP SEQUENCE FROM N.A.
RA Viquez O.M., Konan K.N., Dodo H.W.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF510854; RA046958.1; -
DR HSSP; P04776; IUCX.
DR GO; GO:0045735; F:nutrient reservoir activity; IEA.
DR InterPro; IPR006045; Cupin.
DR InterPro; IPR007113; Cupin region.
DR InterPro; IPR011051; RmlC like cupin.
DR InterPro; IPR006044; Seedstore_115.
DR Pfam; PF00190; Cupin; 2.
DR PRINTS; PR00439; 11SGLBOLIN.
DR PROSITE; PS00305; 11S SEED STORAGE; UNKNOWN 1.
SQ SEQUENCE 538 AA; 61737 MW; 7AABD0D59429709E CRC64;

Alignment Scores:
Pred. No.: 6,04e-163 Length: 538
Score: 2444.00 Matches: 471
Percent Similarity: 93.63% Conservative: 14
Best Local Similarity: 90.93% Mismatches: 19
Query Match: 90.82% Indels: 14
DB: 2 Gaps: 4

US-10-728-051-3 (1-1524) x Q8LKN1 (1-538)
QY 1 CGCGACAAACCGGAGAGAACCGTCCAGTTCAGCGCTCAATCGCAGACCTGAC 60
Db 24 ArgGlnGlnProGluGluAsnAlaCysGlnPheGlnArgLeuAsnAlaGlnArgProAsp 43
QY 61 AATCGCATTTGAATACAGAGCGGCTTACATGAGACTTGGAAACCCCAACACGAGGATTC 120
Db 44 AsnArgIleGluSerGluGlyGlyTyrIleGluThrTrpAsnProAsnAsnGlnGluPhe 63
QY 121 GAATGCGCGGCTCGCCCTCTCTCGTTAGTCTCGCGGCAACCGCCCTTCGTAGCCT 180
Db 64 GluCysAlaGlyValAlaLeuSerArgLeuValLeuArgAsnAlaLeuArgPro 83
QY 181 TTCTACTCCAAATGCTCCCGAGGAGATCTTCATCCAGCAAGGAGGAGTACTTTGGTTG 240
Db 84 PheTyrSerAsnAlaProGlnGluIlePheIleGlnGlnGlyArgGlyTyrPheGlyLeu 103
QY 241 ATATTCCTCGTGTCTCTAGACACTATGAAGAGCCTCACACACAGGTCTCGATCTCAG 300

```
Db 104 IlePheProGlyCysProSerThrTyrGluGluProAlaGlnGlnGlyArgArgHisGln 123
QY 301 TCCCAAGACCCACCAAGAGCTCTCCAAAGGAGAGACCAAAAGGACCAACAGCAACAGATAGT 360
Db 124 SerGlnArgProProArgArgPheGlnGlyGlnAspGlnSerGlnGlnGlnGlnAspSer 143
QY 361 CACCAGAAGTGCACCGTTTCGATGAGGTGATCTCATTTGCAGTTCCCAACCGGTGTGCT 420
Db 144 HisGlnLysValHisArgPheAspGluGlyAspLeuIleAlaValProThrGlyValAla 163
QY 421 TTCTGGCTCTCAACACGACACGACACTGATGTTGTTGCTGTTCTCTTACTGACACCAAC 480
Db 164 PheTrpMetCysAsnAspHisAspThrAspValValAlaValSerLeuThrAspThrAsn 183
QY 481 AACACGCAACACCGAGTTGATCAGTTCCCGCAGAGATTCAATTTGCTGGGAACACGGAG 540
Db 184 AsnAsnAspAsnGlnLeuAspGlnPheProArgArgPheAsnLeuAlaGlyAsnHisGlu 203
QY 541 CAAGATCTTTAAGTACCAGCAACAAAGCAGACAAAGCAGACGAAAGAGCTTACCATAT 600
Db 204 GlnGluPheLeuArgTyrGlnGln-----GlnSerArgArgSerLeuProTyr 220
QY 601 AGCCCATACAGCCGCAAGTCACTAGACAGAGCGGTCAATTTAGCCCTCGAGGA 660
Db 221 SerProTyrSerProGlnThrGlnProLysGlnGluAspArgGluPheSerProArgGly 240
QY 661 CAGCAGACCGCAGAGACGAGCAGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
Db 241 GlnHisGlyArgArgGluArgAlaGlyGlnGluGlnGluAsnGluGlyAsnIlePhe 260
QY 721 AGCGGCTTACCGCGAGTCTCTGGAAACAAGCTTCCAGGTTGACACAGACAGATAGTG 780
Db 261 SerGlyPheThrProGluPheLeuAlaGlnAlaPheGlnValAspAspArgGlnIleLeu 280
QY 781 CAAACCTTAAGAGCGCAGACCGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
Db 281 GlnAsnLeuArgGlyGlnAsnGluSerAspGluGlnGlyAlaIleValThrValArgGly 300
QY 841 GGCTCTCAGAACTTGTAGCCCGACAGATCCCTGACATCTACAACTCAACCTCAAGCTGGT 885
Db 301 GlyLeuArgIleLeuSerProAspArgLysArgArgGlnGlnTyrGluArgProAspGlu 320
QY 886 GAAGAGGATACGATGAAGATGAATATGATGATGATGATGATGATGATGATGATGATGATG 933
Db 321 GluGluGluTyrAspGluAspGluTyrGluTyrAspGluGluGluArgGlnGlnAspArg 340
QY 934 AGCGGTGGCAGGGGAAGCAGAGCGAGGGGAGTGTATTGAAGAGACGATCTGCACCGCA 993
Db 341 ArgArgGlyArgGlySerArgGlySerGlyAsnGlyIleGluGluThrIleCysThrAla 360
QY 994 AGTGCTAAAGAACATTTGGTAGAAACAGATCCCTGACATCTACAACTCAACCTCAAGCTGGT 1053
Db 361 SerPheLysLysAsnIleGlyArgAsnArgSerProAspIleTyrAsnProGlnAlaGly 380
QY 1054 TCACTCAAACTCCCAACGAT-----CTCAACTCTTAATCTAGGTGGCTTGACCT 1107
Db 381 SerLeuLysThrAlaAsnGluLeuGlnLeuAsnLeuIleLeuArgTrpLeuGlyLeu 400
QY 1108 AGTGCTGAATATGAAATCTTACAGGAATGCAATGTTGTCGCTCACTACACACCAAC 1167
Db 401 SerAlaGluTyrGlyAsnLeuTyrArgAsnAlaLeuPheValProHisTyrAsnThrAsn 420
QY 1168 GCACACAGCATCATATATCATTTAGGGAGCGGGCTCAGCTGCAAGTCTGGACAGCAAC 1227
Db 421 AlaHisSerIleIleTyrAlaLeuArgGlyArgAlaHisValGlnValValAspSerAsn 440
QY 1228 GGCACAGAGTGTACAGCAGGAGCTTCAAGAGGTCACGTGTTGTGGTCCACAGAAC 1287
Db 441 GlyAspArgValPheAspGluGluLeuGlnGlnGlyHisValLeuValValProGlnAsn 460
QY 1288 TTCGCGCTCGCTGGAAGTCCCGAGCGAGAACTTCGAAATACGTGGCATTCAGACAGAC 1347
```

```
Db 461 PheAlaValAlaGlyLysSerGlnSerGluAsnPheGluTyrValAlaPheLysThrAsp 480
QY 1348 TCAAGCCGACGATAGCCAACTCCCGGTGAAACTCCGTATATAGATGAACCTCGCGGAG 1407
Db 481 SerArgProSerIleAlaAsnLeuAlaGlyGluAsnSerPheIleAspAsnLeuProGlu 500
QY 1408 GAGGTGGTTGCAAAATTCATATGCGCTCCAAAGGGAGCAGGCAAGCAGCTTAAGAACAC 1467
Db 501 GluValValAlaAsnSerTyrGlyLeuProArgGluGlnAlaArgGlnLeuLysAsnAsn 520
QY 1468 AACCCCTTCAAGTCTTCTGTTCCACCGTCTCAGCAGTCTCCAGGGCTGTGGCT 1521
Db 521 AsnProPheLysPhePheValProProSerGluGlnSerLeuArgAlaValAla 538
RESULT 6
ID Q6T2T4 PRELIMINARY; PRT; 536 AA.
AC Q6T2T4;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Storage protein.
OS Arachis hypogaea (Peanut).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;
OC Arachis.
OX NCBI_TaxID=3818;
RN [1]
RP SEQUENCE FROM N.A.
RA Yang H.-X., Wang F., Bi Y.-P.;
RL Submitted (Oct-2003) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AY439332; AAR02860.1; -.
DR GO; GO:0045735; F:nutrient reservoir activity; IEA.
DR InterPro; IPR006045; Cupin.
DR InterPro; IPR007113; Cupin region.
DR InterPro; IPR011051; RmC_Like_cupin.
DR InterPro; IPR006044; Seedstore_lls.
DR Pfam; PF00190; Cupin; 2.
DR PRINTS; PR00439; 11SGLOBULIN.
DR PROSITE; PS00305; 11S SEED STORAGE; UNKOWN 1.
SQ SEQUENCE 536 AA; 61498 MW; 138F3C881BC2C7D5 CRC64;
Alignment Scores:
Pred. No.: 2,2e-162 Length: 536
Score: 2436.00 Matches: 468
Percent Similarity: 93.60% Conservative: 15
Best Local Similarity: 90.70% Mismatches: 21
Query Match: 90.52% Indels: 12
DB: 2 Gaps: 3
US-10-728-051-3 (1-1524) x Q6T2T4 (1-536)
QY 1 CGGCAGCAACCGGAGAGAACCGTGCACAGTCTCCAGCGCTCAATCGCGAGAGACCTGAC 60
Db 24 ArgGlnGlnProGluGluAsnAlaCysGlnPheGlnArgLeuAsnAlaGlnArgProAsp 43
QY 61 AATCGCATTAATACAGAGCGGTTTACATTGAGACTTGGAAACCCCAACCAACAGAGATTTC 120
Db 44 AsnArgIleGluSerGluGlyGlyTyrIleGluThrTrpAsnProAsnAsnGlnGluPhe 63
QY 121 GAATGGCGGGGTGCGCTCTCTCGTTAGTCTCCCGCGCAACGCCCTTGTAGGCT 180
Db 64 GluCysAlaGlyValAlaLeuSerArgLeuValLeuArgAsnAlaLeuArgArgPro 83
QY 181 TTCTACTCCAACTGCTCCCGAGAGATCTTCATCCAGCAAGGAGGATCTTTGGGTG 240
Db 84 PheTyrSerAsnAlaProGlnGluLeuPheIleGlnGlnGlyArgAlaTyrPheGlyLeu 103
QY 241 ATATTCCCTGGTGTCTTAGACACTATGAAGAGCTTCACACAAAGTCTGTCGATCTCAG 300
Db 104 IlePheLeuGlyCysProSerThrTyrGluGluProAlaGlnGlnGlyArgArgHisGln 123
```


301 TCCCAAGACACCAAGACGCTCCAGGAGAGACCAAGCCACAGCAACGAGTAGT 360
Db SerGlnArgProArgArgPheGlnGlyGlnSerGlnGlnGlnGlnGlnGlnSer 143
361 CACCAAGGTGACACCGCTTCGATGAGGGTGATCTCATTCAGCTTCCACCGGTGTGCT 420
Db HisGlnIysValHisArgPheAspGluGlyAspLeuIleAlaValProThrGlyValAla 163
421 TTCTGGCTCTACCAACGACGACGACGTGATGTTGTTGCTGTTCTTCTTACTGACACCAAC 480
Db PheTrpMetTyrAsnAspHisAspThrAspValValAlaValSerLeuThrAspThrAsn 183
481 AACCAACACACACGCTTGATTCCTCCAGGAGATTCAATTTGGCTGGGACACGGAG 540
Db AsnAsnAspAsnGlnLeuAspGlnPheProArgPheAsnLeuAlaGlyAsnHisGlu 203
541 CAAGAGTCTTAAAGTACCAACGACCAACAAAGCAGACAAAGCAGACGAGAGCTTACCATAT 600
Db GlnGluPheLeuArgTyrGlnGln-----GlnSerArgArgSerLeuProTyr 220
601 AGCCCATACAGCCGCGAAAGTACGCTAGACAGAGAGCGTGAATTTAGCCCTCGAGGA 660
Db SerProTyrSerProGlnThrGlnProIysGlnGluAspArgGluPheSerProArgGly 240
661 CAGCAGACCCGACAGACGACGAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
Db GlnHisGlyArgArgGluArgAlaGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260
721 AGCGGCTTACCGCGGAGTTCCTCGAAACAGACCTCCAGGCTTACAGACAGACAGATAGT 780
Db SerGlyPheThrProGluPheLeuAlaGlnAlaPheGlnValAspAspArgGlnIleLeu 280
781 CAAACCTTAAGAGCGGAGACCGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
Db GlnAsnLeuArgGlyGluAsnGluSerAspGluGlnGlyAlaIleValThrValArgGly 300
841 GGCCTCAGATCTTGACCCAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 885
Db GlyLeuArgIleLeuSerProAspArgLysArgGlnGlnTyrGluArgProAspGlu 320
886 GAAGAGGAATACGATGAAGATGAATATGAATACGATGAAGAGAGAGAGAGAGAGAGAGAGAG 933
Db GluGluGluTyrAspGluAspGluTyrGluTyrAspGluGluGluGluGluGluGluGluGlu 340
934 AGCGTGGCAGGGGAGACAGAGGAGGGGGAATGGTATTAAGAGAGAGAGAGAGAGAGAGAGAG 993
Db ArgArgGlyArgGlySerArgGlySerGlyAsnGlyIleGluGluThrIleCysThrAla 360
994 AGTGCTAAAGAACATTTGCTAGAACAGATCCCTGACATCTACACCTCAGCTGCT 1053
Db SerPheIysLysAsnIleGlyArgAsnArgSerProAspIleTyrAsnProGlnAlaGly 380
1054 TCACTCAAAATGCTCAACGATCTCAACTTCTAATCTTAGTGGCTTGGACCTAGTGCT 1113
Db SerLeuIysThrAlaAsnGluLeuAsnLeuIleLeuArgTyrLeuGlyLeuSerAla 400
1114 GAATATGGAATCTTACAGGAATGCAATTTGCTGCTCCTACACCAACACGACAC 1173
Db GluTyrGlyAsnLeuTyrArgAsnAlaLeuPheValProHisTyrAsnThrAsnAlaHis 420
1174 AGCATCATATATCGATTGAGGAGCGGCTCAGCTGCAAGTCCGTGGACAGACAGCGCAAC 1233
Db SerIleIleTyrAlaLeuArgGlyArgAlaHisValGlnValValAspSerAsnGlyAsp 440
1234 AGAGTGACACGAGGAGCTTCAAGAGGGTCACTGCTGCTGGTGGCCACAGAACTTCGCC 1293
Db ArgValPheAspGluGluLeuGlnGlnGlyHisValLeuValValProGlnAsnPheAla 460
1294 GTGCTCGGAAGTCCAGAGCGGAGAACTTCAATACGTGGCATTCAGACAGACTCAAGG 1353
Db ValAlaGlyLysSerGlnSerGluAsnPheGluTyrValAlaPheIysThrAspSerArg 480
1354 CCAGCATACCAACCTCGCGGGTGAATACTCGTCTATAGTAACCTCGCGGAGGAGTG 1413

Db 481 ProSerIleAlaAsnLeuAlaGlyGluAsnSerPheIleAsnLeuProGluGluVal 500
QY 1414 GTTGCATAATTCATATGCTCCAAAGGAGGAGCAGGCTTAAAGAACAAACCC 1473
Db 501 ValAlaAsnSerTyrGlyLeuProArgGluGlnAlaArgGlnLeuIysAsnAsnPro 520
QY 1474 TTCAGTCTTCTGTTCCAGCGTCTCAGCAGTCTCCGAGGGCTGTGGCT 1521
Db 521 PheLysPheValProSerGluGlnSerLeuArgAlaValAla 536
RESULT 7
Q9SQH7 PRELIMINARY; PRT: 530 AA.
ID Q9SQH7
AC Q9SQH7
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Glycinin.
GN Name=Arachis;
OS Arachis hypogaea (Peanut).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;
OC Arachis.
OX NCBI_TaxID=3818;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Virginia;
RX MEDLINE=99406463; PubMed=10474031; DOI=10.1159/000024203;
RA Kleber-Janke T., Cramer R., Appenzeller U., Schlaak M., Becker W.M.;
RT "Selective cloning of peanut allergens, including profilin and 2S
Int. Arch. Allergy Immunol. 119:265-274 (1999).
RL EMBL; AF086821; AAD47382.1; --
DR HSSP; P04776; IUCX.
DR GO; GO:0045735; F:nutrient reservoir activity; IEA.
DR InterPro; IPR006045; Cupin.
DR InterPro; IPR007113; Cupin_region.
DR InterPro; IPR011051; RMLC_like_cupin.
DR InterPro; IPR006044; Seedstore_11s.
DR Pfam; PF00190; Cupin; 2.
DR PRINTS; PR00439; 11SGLOBULIN.
SQ SEQUENCE 530 AA; 61011 MW; 864E349A31642A32 CRC64;
Alignment Scores:
Pred. No.: 2, 48e-161 Length: 530
Score: 2421.00 Matches: 469
Percent Similarity: 94.09% Conservative: 9
Best Local Similarity: 92.32% Mismatches: 29
Query Match: 89.97% Indels: 2
DB: 2 Gaps: 0
US-10-728-051-3 (1-1524) x Q9SQH7 (1-530)
QY 1 CGGCACCAACCGAGAGAGAACGGCTCCAGCTTCCAGCGCTCAATGCGCGAGACCTGAC 60
Db 24 ArgGlnGlnProGluGluAsnAlaCysGlnPheGlnArgLeuAsnAlaGlnArgProAsp 43
QY 61 AATCGCATTAATCAGAGGCGGTATCATTGAGACTTGAACCCCAACACAGGAGTTTC 120
Db 44 AsnArgIleLeuSerGluGlyGlyIleGluThrTrpAsnProAsnGlnGluPhe 63
QY 121 GAATGCGCGGCTCGCCCTCTCTCGCTTAGTCTCCCGCGCAACGCCCTTCGTAGGCCT 180
Db 64 GluCysAlaGlyValAlaLeuSerArgLeuValLeuArgAsnAlaLeuArgPro 83
QY 181 TTCTACTCAATGCTCCCGAGAGATCTTCATCCAGCAAGAGAGGATATCTTTGGGTTG 240
Db 84 PheTyrSerAsnAlaProGlnGluIlePheIleGlnGlnGlyArgGlyTyrPheGlyLeu 103
QY 241 ATATTCCCTGGTGTCTCCTAGACACTATGAGAGCTTCACACACAAGTCTGCTCATCTCAG 300

```
Db 104 IlePheProGlyCysProSerThrTyrGluGluProAlaGlnGlnGlyArgArgTyrGln 123
QY 301 TCCCAAGACACCAAGACCTCCAGAGGAAGACCAAGCCAAAGCAGCAACGAGATAGT 360
Db 124 SerGlnArgProProArgArgLeuGlnGluGlnSerGlnGlnGlnGlnGlnGlnGln 143
QY 361 CACCAGAGGTGACCGGTTCGATGAGGTGATCTCATTCAGTTCCTCCACCGGTGTGCT 420
Db 144 HisGlnLysValHisArgPheAsnGluGlyAspLeuIleAlaValProThrGlyValAla 163
QY 421 TTCTGGCTCTACAGACACGACACTGATGTGTGCTGCTGCTTCTTACTGACACCAAC 480
Db 164 PheTrpLeuTyrAsnAspHisAspThrAspValValAlaValSerLeuThrAspThrAsn 183
QY 481 AACAAAGACACACGCTTGATCAGTTCCTCCAGAGATTCATTTGCTGGACACGGAG 540
Db 184 AsnAsnAspAsnGlnLeuAspGlnPheProArgPheAsnLeuAlaGlyAsnHisGlu 203
QY 541 CAAGAGTTCCTTAAGGTACCAAGCAACAAAGCAGACGAGCAAGAAAGCTTACCATAT 600
Db 204 GlnGluPheLeuArgTyrGlnGlnSerArgGlnSerArgArgSerLeuProTyr 223
QY 601 AGCCCATACAGCCGCGAAAGTACCTAGCAAGAGAGAGCGTGAATTTAGCCCTCGAGGA 660
Db 224 SerProTyrSerProHisSerArgProArgArgGluGluGluPheArgProArgGly 243
QY 661 CAGCACAGCCGACAGACAGCAGCAGCAAGAGAAAGAAACGAAGTGGAAACATCTTC 720
Db 244 GlnHisSerArgArgGluArgAlaGlyGlnGluGluGluGluGluGluGluGluGlu 263
QY 721 AGCGGTTCACGCGGAGTTCCTGGAAACAGCCCTCCAGGTTGACGACAGACAGATAGTG 780
Db 264 SerGlyPheThrProGluPheLeuGluGlnAlaPheGlnValAspAspArgGlnIleVal 283
QY 781 CAAAACCTAAGAGCGAGACCGAGAGTGAAGAGAGGAGCCATTGTGACAGTGGGGA 840
Db 284 GlnAsnLeuTrpGlyGluAsnGluSerGluGluGluGluGluGluGluGluGluGlu 303
QY 841 GGCTCAGAACTCTGAGCCAGATAGAAAGAGAGCGTCCGACCAAGAAAGAGGAATACGAT 900
Db 304 GlyLeuArgIleLeuSerProAspGlyThrArgGlyAlaAspGluGluGluTyrAsp 323
QY 901 GAAGATGAATATGAATACGATGAAGAGGATAGAAGCGGTGGCAGGGGAAGCAGAGCAGG 960
Db 324 GluAspGlnTyrGluTyrHisGluGlnAspGlyArgArgGlyArgGlySerArgGlyGly 343
QY 961 GGAATGGTATTGAAGAGAGCATCTCCACCGCAAGTGCTAAAGAAACATTGGTAGAAC 1020
Db 344 GlyAsnGlyIleGluGluThrIleCysThrAlaCysValLysLysAsnIleGlyGlyAsn 363
QY 1021 AGATCCCTGACATCTACAACTCTCAAGC-TGGTTCACTCAAACTGCCAAACGATCTCAA 1079
Db 364 ArgSerProHisIleTyrAspProGlnArgTrpPheThrGlnAsnCys-HisAspLeuAs 383
QY 1080 CTTCTAACTACTAGTGGCTTGGACCTAGTGTGTAATATGGAATCTCTACAGGAATCG 1139
Db 383 nLeuLeuIleLeuArgTrpLeuGlyLeuSerAlaGluTyrGlyAsnLeuTyrArgAsnAl 403
QY 1140 ATTGTTGTGCTCACTACAAACCAACGACACGACATCATATATCGATTGAGGGGAG 1199
Db 403 aLeuPheValProHisTyrAsnThrAsnAlaHisSerIleIleTyrAlaLeuArgGlyAr 423
QY 1200 GGCTCAGCTGCAAGTGGTGACAGCAACCGCAACAGAGGTGACGAGGAGCTTCAAGA 1259
Db 423 gAlaHisValGlnValValAspSerAsnGlyAsnArgValTyrAspGluGluLeuGlnG 443
QY 1260 GGTCACGTGCTTGTGGTCCACAGAACTTCGCGCTCGCTGGAAAGTCCAGAGCGAGAA 1319
Db 443 uGlyHisValLeuValValProGlnAsnPheAlaValAlaGlyLysSerGlnSerGluAs 463
QY 1320 CTTTCGAATACGTGGCATTCAGACAGACTCAAGGCCGACGATAGCAACTCGCGGTGA 1379
Db 463 nPheGluTyrValAlaPheLysThrAspSerArgProSerIleAlaAsnPheAlaGlyG 483
```

```
QY 1380 AAATCCGCTCATAGATAACCTCCGAGGAGGTGTTCGAAATTCATATGGCTCCAAAG 1439
Db 483 uAsnSerPheIleAspAsnLeuProGluGluValValAlaAsnSerTyrGlyLeuProAr 503
QY 1440 GGAGCAGGCAAGCGAGCTTAAGACAACAACCCCTTCAAGTTCTTCGTTCCACCGTCTCA 1499
Db 503 gGluGlnAlaArgGlnLeuLysAsnAsnAsnProPheLysPhePheValProPheG 523
QY 1500 GCAGTCTCCGAGGGCTGTGGCT 1521
Db 523 ngInSerProArgAlaValAla 530
RESULT 8
Q61WG5 PRELIMINARY; PRT; 510 AA.
AC Q61WG5;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Glycinin (Fragment)
OS Arachis hypogaea (Peanut).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;
OC Arachis.
OX NCBI_TaxID=3818;
RN [1]_TaxID=3818;
RP SEQUENCE FROM N.A.
RA Kang I.-H., Gallo-Meagher M.;
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY618460; AAT39430.1; -.
DR GO; GO:0045735; P.nutrient reservoir activity; IEA.
DR InterPro; IPR006045; Cupin.
DR InterPro; IPR011051; RmlC_like_cupin.
DR InterPro; IPR006044; Seedstore_11s.
DR Pfam; PF00190; Cupin; 2.
DR PRINTS; PR00439; 11SGLOBULIN.
DR PROSITE; PS00305; 11S_SEED_STORAGE; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 510 AA; 58061 MW; F6569EF13D1A12BD CRC64;
Alignment Scores:
Pred. No.: 3,778-120 Length: 510
Score: 1834.00 Matches: 364
Percent Similarity: 78.57% Conservative: 43
Best Local Similarity: 70.27% Mismatches: 71
Query Match: 68.15% Indels: 40
DB: 2 Gaps: 8
US-10-728-051-3 (1-1524) x Q61WG5 (1-510)
QY 1 CGGACGACACCGGAGAGAGACCGGTGCCAGTTCAGCGCTCAATGCCGAGAGACTGAC 60
Db 22 ArgGlnGlyGlyGluGluAsnGluCysGlnPheGlnArgLeuAsnAlaGlnArgProAsp 41
QY 61 AATCCGATTGAATCAGAGCGGTTCACATTGAGACTTCGAAACCCCAACCAACGAGGTTC 120
Db 42 AsnArgIleGluSerGlnGlyGlyTyrIleGluThrTrpAsnProAsnAsnGlnGluPhe 61
QY 121 GAATCGCGCGGTCCGCTCTCTCGTTCCTAGTTCCTCCCGCCGCAACGCCCTTCGTAGGCT 180
Db 62 GlnCysAlaGlyValAlaLeuSerArgThrValLeuArgAsnAlaLeuArgArgPro 81
QY 181 TTCTACTCCATGCTCCCGAGAGATCTTCATCCAGAGAGGAGGGGATCTTCTGGGTG 240
Db 82 PheTyrSerAsnAlaProLeuGluIleTyrValGlnGlnGlySerGlyTyrPheGlyLeu 101
QY 241 ATATTCCCTGGTGTGCTTAGACATATCAAGAGCGCTCACACACAGGTCTCGATCTCAG 300
Db 102 IlePheProGlyCysProSerThrTyrGluGluProAlaGlnGluGlyArgArgTyrGln 121
QY 301 TCCCAAGACCAACCAAGACGTCTCCAA-----GGAGAGACCAAGCCCAACAGCAACGA 354
```

```
Db 122 SerGlnLysProSerArgArgPheGlnValGlyGlnAspAspProSerGlnGlnGlnGln 141
QY 355 GATAGTCACCAAGAGGTGACCGCTTCGATGAGGGTGATCTCATTCAGTGTCCACCGGT 414
Db 142 AspSerHisGlnLysValHisArgPheAspGluGlyAspLeuIleAlaValProThrGly 161
QY 415 GTTGTCTTCTGGCTCTACAGCAGCAGCAGCAGTGTGTTGCTGTTCTTCTTACTGAC 474
Db 162 ValAlaPheThrMetTyrAsnAspGluAspThrAspValValThrValThrLeuSerAsp 181
QY 475 ACCAAACAACAACGACACGAGCTTGATCAGTTCCTCCAGGAGATTCAATTTGGCTGGGAAC 534
Db 182 ThrSerSerIleHisGlnLysGlnLeuAspGlnPheProArgArgPheTyrLeuAlaGlyAsn 201
QY 535 ACCGAGCAGAGATTCTTAAGGTACACGACCAACAAAGCAGACAAAGCAGACGAAAGCTTA 594
Db 202 GlnGluGlnGluPheLeuArgTyrGlnGlnGlnGlySer----- 215
QY 595 CCATATAGCCCATACAGCCCGCAAGTGCAGCTTACAGACAGAGAGCGTGAATTTAGCCCT 654
Db 216 -----ArgProHisTyrArgGlnIleSerPro 224
QY 655 CGAGGACAGCACACGCCGCGAGAACGAGCAGCAGACGAGCAAGAAACGAGGTGGAAC 714
Db 225 ArgValArg-----GlyAspGlnGlnGlnAsnGluGlySerAsn 237
QY 715 ATCTTCAGCGGCTTCACGCGGAGTTCCTGGAAACAAGCCCTTCAGGTTGACAGCAGACAG 774
Db 238 IlePheSerGlyPheAlaGlnGluPheLeuGlnHisAlaPheGlnVal---AspArgGln 256
QY 775 ATAGTCAAAACCTAGAGCGGAGCAGCAGAGAGTGAAGAGAGGAGGCCATTGTGACAGTG 834
Db 257 ThrValGluAsnLeuArgGlyGluAsnGluArgGluGlnGlyAlaIleValThrVal 276
QY 835 AGGGAGGCTCAGATCTTGACCCAGATAGAAAGAGAGCGTGCC----- 879
Db 277 LysGlyGlyLeuArgIleLeuSerProAspGluAspGluSerArgSerArgSerProPro 296
QY 880 GACCAAGAAGAGGAATACGATGAAGATGAATATATACGATGAAGAGGATGAAGCGGT 939
Db 297 SerArgArgGluGluPheAspGluAspArg-----SerArgProGlnGlnArg 312
QY 940 GGCAGG-----GGAACGACAGGCGAGGGGAATGGTATTGAAGAGCAGATCTGCACC 990
Db 313 GlyLysTyrAspGluAsnArgArgGlyTyrLysAsnGlyIleGluGluThrIleCysSer 332
QY 991 GCAAGTCTAAAGACATTTGCTAGAACACAGATCCCTGACATCTACACCCCTCAAGCT 1050
Db 333 AlaSerValLysLysAsnLeuGlyArgSerSerAsnProAspIleTyrAsnProGlnAla 352
QY 1051 GGTTCACCTCAAACTGCCAACGATCTCAACCTTCTAATCTAGGTGGCTTGGACCTAGT 1110
Db 353 GlySerLeuArgSerValAsnGluLeuAspLeuProIleLeuGlyTrpLeuGlyLeuSer 372
QY 1111 GCTGAATATGAATCTCTACAGAAATGCAATTTGCTGCTCCTCACTACACACCAACGCA 1170
Db 373 AlaGlnHisGlyThrIleTyrArgAsnAlaMetPheValProHisTyrThrLeuAsnAla 392
QY 1171 CACAGCATATATCATGATTGAGGAGCGGCTCAGTGCAGTGCAGTCTGGACAGCAGCGC 1230
Db 393 HisThrIleValValAlaLeuAsnGlyArgAlaHisValGlnValValAspSerAsnGly 412
QY 1231 AACAGAGTGTACGACGAGGAGCTTCAAGAGGGTCACTGCTGCTGGTCCACAGAACTTC 1290
Db 413 AsnArgValTyrAspGluLeuGlnGlnGlyHisValLeuValValProGlnAsnPhe 432
QY 1291 GCCGTGCTGGAAGTCCACAGAGCGAAGACTTCGAATACGTGGCATTCACAGACAGACTCA 1350
Db 433 AlaValAlaAlaLysAlaGlnSerGluAsnTyrGluTyrLeuAlaPheLysThrAspSer 452
QY 1351 AGGCCAGCATAGCAACCTCGCGGTGAAACTCCGTCATAGATAACCTCGCGGAGGAG 1410
```

```
Db 453 ArgProSerIleAlaAsnLeuAlaGlyGluAsnSerIleIleAspAsnLeuProGluGlu 472
QY 1411 GTGGTTGCAAAATTCATATGCGCTCCAAAGGAGCAGCAGGAGGAGTTCAGCAACAAC 1470
Db 473 ValValAlaAsnSerTyrArgLeuProArgGluGlnAlaArgGlnLeuLysAsnAsn 492
QY 1471 CCCTTCAAGTTCTTCGTTCCACCG---TCTCAGCAGTCTCCGAGGCGCTGTGGCT 1521
Db 493 ProPheLysPhePheValProPheAspHisGlnSerMetArgGluValAla 510

RESULT 9
Q647H2 PRELIMINARY; PRT; 484 AA.
ID Q647H2;
AC Q647H2;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Arachin Ahv-3.
OS Arachis hypogaea (Peanut).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;
OC Arachis.
OX NCBI_TaxID=3818;
RN [1]_TaxID=3818;
RP SEQUENCE FROM N.A.
RA Van Y., Wang L., Huang S.;
RT "cDNA clone of peanut seed storage protein gene.";
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AV722687; AAU21492.1; -.
SQ SEQUENCE 484 AA; 54568 MW; 5A3E950752E89D2D CRC64;

Alignment Scores:
Pred. No.: 2,58e-107 Length: 484
Score: 1651.00 Matches: 337
Percent Similarity: 75.25% Conservative: 46
Best Local Similarity: 66.21% Mismatches: 76
Query Match: 61.35% Indels: 50
DB: 2 Gaps: 8

US-10-728-051-3 (1-1524) x Q647H2 (1-484)
QY 1 CGGCAGCACCAGGAGAGAACCGGTGCCAGTTCACGCGCTCAATCGCAGAGACCTGAC 60
Db 24 ArgGlnGlnGlyGluGluAsnGluCysGlnPheGlnArgLeuAsnAlaGlnArgProAsp 43
QY 61 AATCGCATTTGAATCAGAGGCGGTTACATTGAGACTTGGAAACCCCAACACACAGGAGTTC 120
Db 44 AsnCysIleGluSerGluGlyIleGluThrTrpAsnProAsnAsnGlnGluPhe 63
QY 121 GAATGCGCGCGGTGCGCCCTCTCTCGCTTAGTCTCTCCGCGCAACGCCCTTCGTAGCCT 180
Db 64 GlnCysAlaGlyValAlaLeuSerArgPheValLeuArgAsnAlaLeuArgPro 83
QY 181 TTCTACTCAATGCTCCCGAGGAGATCTTCATCCAGCAAGGAGGAGTACTTTGGGTG 240
Db 84 PheTyrSerAsnAlaProGlnGluIlePheIleTyrGlnGlySerGlyTyrPheGlyLeu 103
QY 241 ATATTCCTCGGTGTCTAGACACTATGAAGAGCCTCACACAAAGTCTGTCGATCTCAG 300
Db 104 IlePheProGlyCysProGlyThrPheGluGluPro---IleGlnGly---SerGluGln 121
QY 301 TCCCAAGACCAACAGAGCGTCTCCAAAGGAGAGACCAAGCCCAACAGCAGATAGT 360
Db 122 PheGlnArgProSerArgHisPheGlnGlyGlnAspGlnSerGlnArgProLeuAspThr 141
QY 361 CACGAGAGGTGACCGTTTCGATGAGGTGATCTCATTCAGTTCCTCCACCGGTGTGCT 420
Db 142 HisGlnLysValHisGlyPheArgGluGlyAspLeuIleAlaValProHisGlyValAla 161
QY 421 TTCTGCTCTACACCAACGACGACACTGATGTTGTTGCTGTTTCTCTTACTGACACCAAC 480
Db 162 PheTrpIleTyrAsnAspGlnAspThrAspValValAlaIleSerValLeuHisThrAsn 181
```

```

QY 481 AACACGACACACGAGCTTGATGATGCTTCCCGCAGGAGATTCATTTGGCTGGGAACACGGAG 540
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 182 SerLeuHisAsnGlnLeuAspGlnPheProArgArgPheAsnLeuAlaGlyLysGlnGlu 201
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 541 CAAGAGTCTTAAGGTACGACCAACAAAGC---AGACAAAGCAGACGAGAACGCTTACCA 597
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 202 GlnGluPheLeuArgfyrGlnGlnArgSerGlyArgGlnSer----- 215
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 598 TATAGCCCATACAGCCCGCAAGTGCAGCTAGACAAGAGAGCGTGATTTAGCCCTCGA 657
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 216 ----- 222
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 658 GGACAGCAGCCGACAGAACGAGCAGGACAGAAAGAAACCAAGAGTGGAAACATC 717
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 223 -----GlnGluGlnGluAsnGluGlyGlyAsnVal 232
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 718 TTCAGCGGCTTACCGCGGAGTTCCTGGAACAGCCTTCAGGTTACGACGACAGATTA 777
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 233 PheSerGlyPheSerThrGluPheLeuSerHisGlyPheGlnValAsnGlu---AspIle 251
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 778 GTCAAAACCTAAGAGCGCAGACCGAGAGTGAAGAGAGAGGAGGCCATTGTGACAGTGAGG 837
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 252 ValArgAsnLeuArgGlyGluAsnGluArgGluGlnGlyAlaIleValThrValLys 271
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 838 GGAGGCTCAGATCTTGAGCCCGACAGATAGAGAGAGAGAGAGCGTCCGACGAGAGGAATAC 897
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 272 GlyGlyLeuSerIleLeuValPro----- 280
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 898 GATGAGATGAATATGAATACGATGAAGAGATAGAGGCGTGGCAGGGAAGCAGCAGGC 957
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 281 -----GluTrpArgGlnSerTyrGlnGlnProGlyArgGlyAspLysAsp 295
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 958 AGGGGAATGTTTGAAGAGAGCGATCTGCACCGCAAGTGCTAAAGAACATTTGTGAGA 1017
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 296 PheAsnAsnGlyIleGluGluThrIleCysThrAlaThrValLysMetAsnIleGlyLys 315
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1018 AACAGATCCCTGACATCTACACCTCAAGCTGTGTTCACTCAAACTGCCAACGATCTC 1077
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 316 SerThrAlaAspIleTyrAsnProGlnAlaGlySerValArgThrValAsnGluLeu 335
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1078 AACCTTCTAATCTAGGTGGCTTGACCTAGCTAGTGAATATGGAATCTTACAGGAAT 1137
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 336 AspLeuProIleLeuAsnArgLeuGlyLeuSerAlaGluTyrGlySerIleHisArgAsp 355
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1138 GCATTTGTTGCTCACTACACCAACGACGACATCATATATCGATTGAGGGA 1197
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 356 AlaMetPheValProHisTyrAsnMetAsnAlaAsnSerMetIleTyrAlaLeuHisGly 375
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1198 CGGGCTCAGTGAAGTGTGGACGACGACGACGAGGTGTACGACGAGGAGCTTCAA 1257
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 376 GlyAlaHisValGlnValValAspCysAsnGlyAsnArgValPheAspGluGluLeuGln 395
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1258 GAGGTCACGTGCTGTGTGTGCACAGAACTTCGCGCTCGCTGGAAGTCCAGAGCGAG 1317
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 396 GluGlyGlnSerLeuValValProGlnAsnPheAlaValAlaLysSerGlnSerGlu 415
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1318 AACTTCGAATACGTGGCTTCAAGACAGACTCAAGGCCCGCAGCATACCACTCGCGGT 1377
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 416 HisPheLeuTyrValAlaPheLysThrAsnSerArgAlaSerIleSerAsnLeuAlaGly 435
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1378 GAAATCTCGTCATAGATACCTGCGGAGGAGGTGTGCAATTCATATGCTCCCAA 1437
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 436 LysAsnSerTyrMetTrpAsnLeuProGluAspValValAlaAsnSerTyrGlyLeuGln 455
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1438 AGGGACGAGCAGGAGCTTAAGACAAACACCCCTTCAAGTCTTCGTTCCACG--- 1494
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 456 TyrGluGlnAlaArgGlnLeuLysAsnAsnProPheThrPheLeuValProGln 475
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1495 TCTACAGCTCTCCGAGGCGTGTGCT 1521
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 476 AspSerGlnMetIleArgThrValAla 484
   : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 10

```

GLC3_SOYBN STANDARD; PRT; 481 AA.
ID GLC3_SOYBN
AC P11828;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Glycinin G3 precursor [Contains: Glycinin A subunit; Glycinin B
subunit].
GN Name=GY3;
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]_TaxID=3847;
SEQUENCE FROM N.A.
RC STRAIN=cv. Dare; TISSUE=Leaf;
RX MEDLINE=89296500; PubMed=2740231;
RA Cho T.-J., Nielsen N.C.;
RT "The glycinin G3 gene from soybean.";
RL Nucleic Acids Res. 17:4388-4388(1989).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=9239391; PubMed=2485233;
RA Nielsen N.C., Dickinson C.D., Cho T.-J., Thanh V.H., Scallan B.J.,
RA Fischer R.L., Sims T.L., Drews G.N., Goldberg R.B.;
RT "Characterization of the glycinin gene family in soybean.";
RL Plant Cell 1:313-328(1989).
CC -!- FUNCTION: Glycinin is the major seed storage protein of soybean.
CC -!- SUBUNIT: Hexamer; each subunit is composed of an acidic and a
CC basic chain derived from a single precursor and linked by a
CC disulfide bond.
CC -!- SIMILARITY: Belongs to the 11S seed storage protein (globulins)
family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
EMBL; X15123; CAA33217.1; -.
DR PIR; S04605; S04605.
DR HSP; P04776; 1EXZ.
DR InterPro; IPR006045; Cupin.
DR InterPro; IPR007113; Cupin_region.
DR InterPro; IPR011051; Rmlc_like_cupin.
DR InterPro; IPR006044; Seedstore_11s.
DR Pfam; PF00190; Cupin; 2.
DR PRINTS; PR00439; 11SGLOBULIN.
DR PROSITE; PS00305; 11S_SEED_STORAGE; 1.
KW Multigene family; Seed storage protein; Signal.
FT SIGNAL 1 21
FT CHAIN 22 296 Glycinin A subunit.
FT CHAIN 297 476 Glycinin B subunit.
FT PROPEP 477 481
FT DISULFID 107 303 Interchain (By similarity).
SQ SEQUENCE 481 AA; 54241 MW; 5F5C3148DF6241A7 CRC64;

```

Alignment Scores:

```

Pred. No.: 7,33e-97 Length: 481
Score: 1502.00 Matches: 297
Percent Similarity: 71.57% Conservative: 68
Best Local Similarity: 58.24% Mismatches: 91
Query Match: 55.82% Indels: 54
DB: 1 Gaps: 7

```

US-10-728-051-3 (1-1524) x GLC3_SOYBN (1-481)

QY 1 CGGCAGCAGCCGAGGAGAAAGCGTCCAGCTTCAGCGCTCAATGCGCAGAGACCTGAC 60

```
Db 23 ArgGluGlnProGlnGlnGluCysGlnIleGlnArgLeuAsnAlaLeuLysProAsp 42
QY 61 AATCGCATTAATCAGAGGCGGTTTACATTGAGACTTGGAAACCCCAACACAGGAGTTC 120
Db 43 AsnArgIleGlnSerGluGlyGlyPheIleGluThrTrpAsnProAsnAsnLysProPhe 62
QY 121 GAATGCGCGCGCTCGCCCTCTCTCGCTTAGTCTCCGCGCGCAACGCCCTCTCGTAGGCCT 180
Db 63 GlnCysAlaGlyValAlaLeuSerArgCysThrLeuAsnArgAsnAlaLeuArgArgPro 82
QY 181 TTCTACTCCAACTCCCGAGGAGATCTTCATCCAGCAAGGAGGAGTACTTTGGGTG 240
Db 83 SerTyrThrAsnAlaProGlnGluIleTyrIleGlnGlySerGlyIlePheGlyMet 102
QY 241 ATATTCCCTCGTGTCTTACACTACTAAGAGCGCTCACACAAAGTCTCGTCAATCTCAG 300
Db 103 IlePheProGlyCysProSerThrPheGluGluProGlnGlnLysGly-----Gln 119
QY 301 TCCCAAGACCAACAAGACGCTCCCAAGGAGAGACAAAGCCAAAGCAACAGCAATAGT 360
Db 120 SerSerArgPro-----GlnAspArg 126
QY 361 CACGAGAGGTGCACCGTTTCGATGAGGTGATCTCATTTGCAGTTCACCGCGGTGTGCT 420
Db 127 HisGlnLysIleTyrHisPheArgGluGlyAspLeuIleAlaValProThrGlyPheAla 146
QY 421 TTCTGGCTTACCAACGACCAACGACACTGATGTTGTGCTGTTCTTCTTACTGACACCAAC 480
Db 147 TyrTrpMetTyrAsnAsnGluAspThrProValValAlaValSerLeuIleAspThrAsn 166
QY 481 AACACGACACACAGCTTGATGATTTCCCGAGGAGATTCAATTTGGCTGGGAACACGGAG 540
Db 167 SerPheGlnAsnGlnLeuAspGlnMetProArgPheTyrLeuAlaGlyAsnGlnGlu 186
QY 541 CAGAGTCTTAGGTACGACGACCAACAGCAGACCAAGCAGACGAGAGAGAGCTTACCATAT 600
Db 187 GlnGluPheLeuGlnTyrGlnProGlnLysGlnGlnGly----- 199
QY 601 AGCCCATACACCGCGCAAGTCCAGCTAGACAGAGAGCGGTGAATTTAGCCCTCGAGGA 660
Db 200 -----Gly 200
QY 661 CAGCAGACCGCCAGAGACGACGAGCAGCAAGAGAAAGAAACGAAGTGGAAACATCTTC 720
Db 201 ThrGlnSerGlnLysGlyLysArgGlnGlnGluGluAsnGluGlyGlySerIleLeu 220
QY 721 AGCGCTTACGCGCGAGTCTCGGAACAAGCCCTCCAGGTTGACGACGACAGATAGTG 780
Db 221 SerGlyPheAlaProGluPheLeuGluHisAlaPheValVal---AspArgGlnIleVal 239
QY 781 CAAACCTAAGAGCGAGACCGAGAGTGAAGAGAGAGGAGCCATTGTGACAGTGAGGGGA 840
Db 240 ArgLysLeuGlnGlyLeuAsnGluGluGluGlyAlaIleValThrValLysGly 259
QY 841 GGCCTCAGAATCTTGACCCCA-----GATAGAAAGAGACGTCGCGACGAAGAAGAG 891
Db 260 GlyLeuSerValIleSerProProThrGluGluGlnGlnArgProGluGluGluGlu 279
QY 892 GATACCATGAAGATGAATATGATGATGAGAGATAGAAGCGCTGCGCAGGGAGAGC 951
Db 280 LysProAsp-----CysAspGluLysAspLysHisCysGlnSerGlnSer 294
QY 952 AGAGCGAGGGGATGTTTGAAGACAGCATCTGCACCGCAAGTCGTAAAGACATT 1011
Db 295 Arg-----AsnGlyIleAspGluThrIleCysThrMetArgLeuArgHisAsnIle 311
QY 1012 GGTAGAACAGATCCCTGATCATACAAACCTCAAGCTGTTCACTCAAAACTGCCAAC 1071
Db 312 GlyGlnThrSerSerProAspIlePheAsnProGlnAlaGlySerIleThrAlaThr 331
QY 1072 GATCTCAACCTTCTAACTACTAGTGGCTTGGACCTAGTGTCTGAATATGGAATCTCTAC 1131
```

```
Db 332 SerLeuAspPheProAlaLeuSerTrpLeuLysLeuSerAlaGlnPheGlySerLeuArg 351
QY 1132 AGAATGCAATTTGTTGCTCTACTACACCAACACGACGACAGCATCATATATCATGTTG 1191
Db 352 LysAsnAlaMetPheValProHisTyrAsnLeuAsnAlaAsnSerIleIleTyrAlaLeu 371
QY 1192 AGGGACCGGCTCAGTCCAGTCTGGACACGCAACGCGCAACAGAGTGACGACGAGGAG 1251
Db 372 AsnGlyArgAlaLeuValGlnValValAsnCysAsnGlyGluArgValPheAspGlyGlu 391
QY 1252 CTTCAAGAGGGTCAAGTCTGTTGCTGTCACAGAACTTCGCGCTCGCTGGAAGTCCAG 1311
Db 392 LeuGlnGluGlyGlnValLeuIleValProGlnAsnPheAlaValAlaArgSerGln 411
QY 1312 ACGAGAACTTCGAATACGTGCTTCAAGACAGACTCAAGCCAGCATAGCCAACTTC 1371
Db 412 SerAspAsnPheGluTyrValSerPheLysThrAsnAspArgProSerIleGlyAsnLeu 431
QY 1372 GCGGTGAAACTCCGTCATAGATAACCTGCGCGGAGGAGTGTTCNAATTCATATGCG 1431
Db 432 AlaGlyAlaAsnSerLeuLeuAlaLeuProGluGluValIleGlnGlnThrPheAsn 451
QY 1432 CTCCAAGGAGGACGAGGCAAGGAGCTTAAAGAACAAACCCCTTCAAGTTCCTGTTCCA 1491
Db 452 LeuArgArgGlnGlnAlaArgGlnValLysAsnAsnAsnProPheSerPheLeuValPro 471
QY 1492 CCGTCTCACAGCTCTCCGAGGCTGTGGCT 1521
Db 472 ProLysGluSerGlnArgValValAla 481
RESULT 11
GLC2_SOYBN STANDARD; PRT; 485 AA.
ID AC P0405; P04121; P04348; P04349;
DT 20-MAR-1987 (Rel. 04, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 05-JUL-2004 (Rel. 14, Last annotation update)
DE Glycinin G2 precursor [Contains: Glycinin A2 subunit; Glycinin B1a subunit].
GN Names=Gly2;
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosids;
OC Euphorbiales I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Dare; TISSUE=Leaf;
RX MEDLINE=92393391; PubMed=2485233;
RA Nielsen N.C., Dickinson C.D., Cho T.J., Thanh V.H., Scallan B.J.,
RA Fischer R.L., Sims T.L., Drews G.N., Goldberg R.B.;
RT "Characterization of the glycinin gene family in soybean.";
RL Plant Cell 1:313-328 (1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Dare; TISSUE=Leaf;
RX MEDLINE=89296499; PubMed=2740230;
RA Thanh V.H., Turner N.E., Nielsen N.C.;
RT "The glycinin Gy2 gene from soybean.";
RL Nucleic Acids Res. 17:4387-4387 (1989).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Shirotaurunoko;
RA Utsumi S., Kim C.S., Kohno M., Kito M.;
RT "Polymorphism and expression of cDNAs encoding glycinin subunits.";
RL Agric. Biol. Chem. 51:3267-3273 (1987).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=88040439; PubMed=3671077;
RA Fukazawa C., Momma T., Higuchi W., Uda K.;
RT "Complete nucleotide sequence of the gene encoding a glycinin A2B1a subunit precursor of soybean.";
RL Nucleic Acids Res. 15:8117-8117 (1987).
```

[5] SEQUENCE FROM N.A.
RA Momma T., Negoro T., Uga K., Fukazawa C.;
RA "A complete cDNA coding for the sequence of glycinin A2B1a subunit
RT precursor.";
RT FEBS Lett. 188:117-122(1985).
RN [6]
RN SEQUENCE OF 19-296 AND 301-480 (A2 AND B1a SUBUNITS).
RP MEDLINE=85030470; PubMed=6541652;
RX Staswick P.E., Hermodson M.A., Nielsen N.C.;
RA "The amino acid sequence of the A2B1a subunit of glycinin.";
RT J. Biol. Chem. 259:13424-13430(1984).
RN [7]
RN SEQUENCE OF 262-485 FROM N.A.
RP MEDLINE=85030472; PubMed=6092376;
RX Marco Y.A., Thanh V.H., Tumer N.E., Scallan B.J., Nielsen N.C.;
RA "Cloning and structural analysis of DNA encoding an A2B1a subunit of
RT glycinin.";
RT J. Biol. Chem. 259:13436-13441(1984).
RN [8]
RN SEQUENCE OF 1-36 FROM N.A.
RP MEDLINE=90332420; PubMed=2377465;
RX Kitamura Y., Arahira M., Itoh Y., Fukazawa C.;
RA "The complete nucleotide sequence of soybean glycinin A2B1a gene
RT spanning to another glycinin gene AlaB1b.";
RT Nucleic Acids Res. 18:4245-4245(1990).
RN [9]
RN DISULFIDE BOND.
RP MEDLINE=85030471; PubMed=6541653;
RX Staswick P.E., Hermodson M.A., Nielsen N.C.;
RA "Identification of the cysteines which link the acidic and basic
RT components of the glycinin subunits.";
RT J. Biol. Chem. 259:13431-13435(1984).
CC -1- FUNCTION: Glycinin is the major seed storage protein of soybean.
CC -1- SUBUNIT: Hexamer; each subunit is composed of an acidic and a
CC basic chain derived from a single precursor and linked by a
CC disulfide bond.
CC
CC -1- SIMILARITY: Belongs to the 11S seed storage protein (globulins)
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X15122; CAA33216.1; -
DR EMBL; D00216; BAA00154.1; -
DR EMBL; Y00398; CAA68460.1; -
DR EMBL; X02806; CAA26575.1; -
DR EMBL; X02646; AAA33963.1; -
DR EMBL; X53404; CAA37480.1; -
DR PIR; A91341; FWSYGI.
DR PIR; S11002; S11002.
DR HGSP; P04776; 1FXZ.
DR InterPro; IPR006045; Cupin.
DR InterPro; IPR011051; RnLc_like_cupin.
DR InterPro; IPR006044; Seedstore_11s.
DR Pfam; PF00190; Cupin; 2.
DR PRINTS; PR00439; 11SGLOBULIN.
DR PROSITE; PS00305; 11S_SEED_STORAGE; 1.
KW Direct protein sequencing; Multigene family; Seed storage protein;
KW Signal.
FT SIGNAL 1 18
FT CHAIN 19 296 Glycinin A2 subunit.
FT PROPEP 297 300
FT CHAIN 301 480 Glycinin B1a subunit.
FT PROPEP 481 485
FT DISULFID 104 307 Interchain.
FT VARIANT 103 103 G -> D.
FT VARIANT 318 318 N -> T.

Db 238 ArgAsnLeuGlnGlyGluAsnGluGluAspSerGlyAlaIleValThrValValGly 257
QY 841 GGCTCAGAACTTGACCCAGATAGAAAGAGAGCGTCCGACGAAAGAGAGGAAATACGAT 900
Db 258 GlyLeuArgValThrAlaProAlaMetArgLysProGlnGlnGluAspAsp 277
QY 901 GAAGATGAA-----TATGATACATGACAGAGATAGAGGCGTGCAGGGGA 948
Db 278 GluGluGluGlnProGlnCysValGluThrAspLysGlyCysGlnArgGlnSerLysArg 297
QY 949 AGCAGAGCGCGGGAATGTTTAAAGACAGCATCTGCACCCCAAGTGTCTAAAGAAAC 1008
Db 298 SerArg-----AsnGlyIleAspGluThrIleCysThrMetArgLeuArgGlnAsn 314
QY 1009 ATTGGTAGAAACAGATCCCTGACATCTACAACCTCAAGCTGGTTCACCTCAAAATGCC 1068
Db 315 IleGlyGlnAsnSerProAspIleTyrAsnProGlnAlaGlySerIleThrAla 334
QY 1069 AACGATCTCAACCTTCTAATCTAGTGGCTGGACCTAGTCTGAATATGGAATCTC 1128
Db 335 ThrSerLeuAspPheProAlaLeuTrpLeuLeuLysLeuSerAlaGlnTyrGlySerLeu 354
QY 1129 TACAGGAATCATTTGTTTGGCTCACTACAACCAACGACACAGCATCATATATCGA 1188
Db 355 ArgIysAsnAlaMetPheValProHisTyrThrLeuAsnAlaAsnSerIleIleTyrAla 374
QY 1189 TTGAGGGGACGGCTCAGTGCAGTCTGCGACGAGCAACGCGACAGAGTGTACGACGAG 1248
Db 375 LeuAsnGlyArgAlaLeuValGlnValAlaAsnCysAsnGlyGluArgValPheAspGly 394
QY 1249 GAGCTTCAAGAGGTACAGTGTGTTGTCGTCACAGAACTTCGCGTCTGCGTGAAGTCC 1308
Db 395 GluLeuGlnGlyGlyValLeuIleValProGlnAsnPheAlaValAlaLysSer 414
QY 1309 CAGAGCGAGAACTTGAATACGTGGCATTCAGACAGACTCAAGCCACATAGCAAC 1368
Db 415 GlnSerAspAsnPheGluTyrValSerPheLysThrAsnAspArgProSerIleGlyAsn 434
QY 1369 CTCGCCGCTGAAACTCCGTCTAGATAAACCTGCGGAGAGGTGGTTCGCAATTCATAT 1428
Db 435 LeuAlaGlyAlaAsnSerLeuLeuAsnAlaLeuProGluValIleGlnHisThrPhe 454
QY 1429 GGCTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1488
Db 455 AsnLeuLysSerGlnGlnAlaArgGlnValLysAsnAsnAsnProPheSerPheLeuVal 474
QY 1489 CCACCGCTCAGCAGTCTCCGAGGGGTGCGCT 1521
Db 475 ProProGlnGluSerGlnArgArgAlaValAla 485

RESULT 12
GLC1_SOYBN STANDARD; PRT; 495 AA.
AC P04776;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Glycinin G1 precursor [Contains: Glycinin ALA subunit; Glycinin BX subunit].
DE Names=Gly;
GN Glycine max (Soybean).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Dare; TISSUE=Leaf;
RX MEDLINE=92393391; PubMed=2485233;
RA Nielsen N.C., Dickinson C.D., Cho T.J., Thanh V.H., Scallion B.J.,
RA Fischer R.L., Sims T.L., Drews G.N., Goldberg R.B.;
RT "Characterization of the glycinin gene family in soybean.";
RL Plant Cell 1:313-328(1989).

[2] SEQUENCE FROM N.A.
RC STRAIN=cv. Dare; TISSUE=Leaf;
RX MEDLINE=89296498; PubMed=2740229;
RA Sims T.L., Goldberg R.B.;
RT "The glycinin Gyl gene from soybean.";
RL Nucleic Acids Res. 17:4386-4386(1989).
[3] SEQUENCE FROM N.A.
RC STRAIN=cv. Boninori;
RX MEDLINE=86041867; PubMed=2997720;
RA Negoro T., Momma T., Fukazawa C.;
RT "A cDNA clone encoding a glycinin Ala subunit precursor of soybean.";
RL Nucleic Acids Res. 13:6719-6731(1985).
[4] SEQUENCE FROM N.A.
RC Utsumi S., Kohno M., Mori T., Kito M.;
RT "An alternate cDNA encoding glycinin Ala Bx subunit.";
RL J. Agric. Food Chem. 35:210-214(1987).
[5] SEQUENCE FROM N.A.
RC Urade R., Nakatani H., Nakano C.;
RT "mRNA of soybean proglycinin AlaB1b subunit.";
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: Glycinin is the major seed storage protein of soybean.
CC -I- SUBUNIT: Hexamer; each subunit is composed of an acidic and a
CC disulfide bond.
CC -I- PTM: The precursor is post-translational processed to form a
CC covalently linked ALA-BX subunit complex.
CC -I- SIMILARITY: Belongs to the 11S seed storage protein (globulins)
CC family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M36686; AAA33966.1; -;
CC EMBL; X15121; CAA33215.1; -;
CC EMBL; X02985; CAA36723.1; -;
CC EMBL; AB113349; BAC78522.1; -;
CC PIR; A23497; FMSY62.
CC PIR; S10851; S10851.
CC PDB; 1FXZ; X-ray; A/B/C=20-495.
CC PDB; 1UCK; X-ray; A/B/C=20-495.
CC PDB; 1UD1; X-ray; A/B/C=20-495.
CC InterPro; IPR006045; Cupin.
CC InterPro; IPR007113; Cupin region.
CC InterPro; IPR011051; RmlC like cupin.
CC InterPro; IPR006044; Seedstore_11s.
CC Pfam; PF00190; Cupin; 2.
CC PRINTS; PR00439; 11SGLOBULIN.
CC PROSITE; PS00305; 11S SEED STORAGE; 1.
KW 3D-structure; Multigene family; Seed storage protein; Signal.
FT SIGNAL 1 19
FT CHAIN 20 306 Glycinin ALA subunit.
FT PROPEP 307 310
FT CHAIN 311 490 Glycinin BX subunit.
FT PROPEP 491 495
FT DISULFID 107 317 Interchain (By similarity).
FT CONFLICT 42 42 D -> G (in Ref. 3).
FT CONFLICT 108 108 P -> S (in Ref. 3).
FT CONFLICT 136 136 E -> S (in Ref. 3).
FT CONFLICT 360 360 E -> G (in Ref. 3).
FT TURN 30 31
FT STRAND 39 39
FT STRAND 43 47
FT TURN 48 48
FT STRAND 50 54

FT	TURN	57	58		
FT	HELIX	60	65		
FT	TURN	66	66		
FT	STRAND	68	74		
FT	TURN	76	77		
FT	STRAND	78	84		
FT	STRAND	89	95		
FT	STRAND	97	103		
FT	TURN	105	106		
FT	STRAND	134	137		
FT	TURN	138	139		
FT	STRAND	140	144		
FT	TURN	146	147		
FT	STRAND	149	154		
FT	STRAND	160	166		
FT	TURN	168	169		
FT	TURN	171	172		
FT	STRAND	181	183		
FT	TURN	191	192		
FT	HELIX	193	196		
FT	TURN	222	223		
FT	HELIX	226	233		
FT	TURN	234	234		
FT	HELIX	237	243		
FT	STRAND	255	257		
FT	TURN	317	318		
FT	STRAND	322	324		
FT	STRAND	333	335		
FT	TURN	336	338		
FT	STRAND	339	344		
FT	TURN	346	348		
FT	HELIX	350	352		
FT	TURN	353	356		
FT	STRAND	359	364		
FT	TURN	366	367		
FT	STRAND	369	375		
FT	TURN	376	377		
FT	STRAND	380	386		
FT	STRAND	388	394		
FT	TURN	396	397		
FT	STRAND	400	407		
FT	TURN	408	409		
FT	STRAND	411	414		
FT	TURN	416	417		
FT	STRAND	419	424		
FT	TURN	427	428		
FT	STRAND	429	435		
FT	STRAND	442	444		
FT	TURN	448	449		
FT	TURN	451	451		
FT	HELIX	452	454		
FT	HELIX	457	464		
FT	TURN	465	465		
FT	HELIX	468	476		
FT	STRAND	483	484		
SQ	SEQUENCE	495 AA;	55706 MW; 329CB0545B24D894 CRC64;		

Alignment Scores:
Pred. No.: 1 94e-96 Length: 495
Score: 1496.00 Matches: 292
Percent Similarity: 70.87% Conservative: 73
Best Local Similarity: 56.70% Mismatches: 100
Query Match: 55.59% Indels: 50
DB: 1 Gaps: 6

US-10-728-051-3 (1-1524) x GLC1_SOYBN (1-495)

QY	1	CGCAGCAACCGGAGGAGAACGGTCCAGTTCACAGCGCTCAATCGCAGAGACCTGAC	60
Db	23	ArgGluGlnProGlnGlnAsnGluCysGlnIleGlnLysLeuAsnAlaLeuLysProAsp	42
QY	61	AATCGCATTAATCAGAGGCGGTACATTGACACTTGGNACCCCAACACAGGAGTTC	120

Db	43	AsnArgIleGluSerGluGlyGlyLeuIleGluThrTrpAsnProAsnAsnLysProPhe	62
QY	121	GAATCGCGCGGTGCGCTCTCTCGCTTAGTCTCCGCCCAAGCCCTTCGTAGGCT	180
Db	63	GlnCysAlaGlyValAlaLeuSerArgCysThrLeuAsnArgAsnAlaLeuArgPro	82
QY	181	TTCTACTCCAATGCTCCCGAGAGATCTTCACAGCAAGAGGAGATCTTTGGGTG	240
Db	83	SerTyrThrAsnGlyProGlnGluIleTyrIleGlnGlnGlyLysGlyIlePheGlyMet	102
QY	241	ATATTCCCTGGTGTCTAGACACTATGAGAGCTCACACAAAGGTGCTCGATCTCAG	300
Db	103	IleTyrProGlyCysProSerThrPheGluGluProGlnGlnProGlnArgGlyGln	122
QY	301	TCCCAAGACCAACAGACGTCTCCAAGAGAGAACCAAGCCAAAGCAGCAACGAGATAGT	360
Db	123	SerSerArgPro-----GlnAspArg	129
QY	361	CACCAAGAGTGCACCGTTTCGATGAGGTGATCTCATTGCAAGTCCCAACCGGTGCT	420
Db	130	HisGlnLysIleTyrAsnPheArgGluGlyAspLeuIleAlaValProThrGlyValAla	149
QY	421	TTCTGGCTCTACACGACCACGACACTGATGTGTGTCTTCTTACTCACACCAAC	480
Db	150	TrpTrpMetTyrAsnAsnGluAspThrProValValAlaValSerIleAspThrAsn	169
QY	481	AACAACGACACACAGCTTGATCAGTTCCCGAGAGATTCAATTTGGTGGGAACACGGAG	540
Db	170	SerLeuGluAsnGlnLeuAspGlnMetProArgArgPheTyrLeuAlaGlyAsnGlnGlu	189
QY	541	CAAGAGTTCTTAAGTACCAAGCAACAAAGCAGACAAAGCAGACGAAAGCTTACCATAT	600
Db	190	GlnGluPheLeuLysTyrGlnGlnGln-----	199
QY	601	AGCCCATACAGCCCGCAAGTCAGCTAGACAAAGAGCGTGAATTTAGCCCTCGAGGA	660
Db	200	-----GlyGly	201
QY	661	CAGCACAGCCGACAGACGAGCAAGAGAAAGAAAGAAAGAGTGGAAACATCTTC	720
Db	202	HisGlnSerGlnLysGlyHisGlnGlnGluGluAsnGlnGlyGlySerIleLeu	221
QY	721	AGCGCTTCACCGCGAGTTCCTCGAACAGCTTCCAGGTTGACGACAGACAGATAGTG	780
Db	222	SerGlyPheThrLeuGluPheLeuGluHisAlaPheSerVal---AspLysGlnIleAla	240
QY	781	CAAAACCTAAGAGCGGACGACCGAGAGTGAAGAGAGGAGGCCATTGTGACAGTGGGGA	840
Db	241	LysAsnLeuGlnGlyGluAsnGluGlyGluAspLysGlyAlaIleValThrValLysGly	260
QY	841	GGCTCAGAAATCTTGAGCCCA-----GATAGAAAGACGTCGCCGACGAAAGAG	891
Db	261	GlyLeuSerValIleLysProProThrAspGluGlnGlnGlnArgProGlnGluGlu	280
QY	892	GAATACGATGAAGATGAATATGAATAC-----GATGAACAG---GATGAAGG	936
Db	281	GluGluGluGluAspLysProGlnCysLysGlyLysAspLysHisCysGlnArgPro	300
QY	937	CGTGGCAGGGAGCAGAGCGCGGGGATGTTAGACGATCTGCACCGCAAGT	996
Db	301	ArgGlySerGlnSerLysSerArgArgAsnGlyIleAspGluThrIleCysThrMetArg	320
QY	997	GCTAAAAGAACATTGGTAGAAACAGATCCCTGACATCTACAACCTCAAGCTGGTCA	1056
Db	321	LeuArgHisAsnIleGlyGlnThrSerSerProAspIleTyrAsnProGlnAlaGlySer	340
QY	1057	CTCAAAATCGCAACGATCTCAACCTTCTAATCTTAGTGGCTGGACCTAGTCTGAA	1116
Db	341	ValThrThrAlaThrSerLeuAspPheProAlaLeuSerTrpLeuArgLeuSerAlaGlu	360
QY	1117	TATGGAATCTCTACAGGAATGATTTGTTGCTGCTACTACAACACCAACCCACACAGC	1176
Db	361	PheGlySerLeuArgLysAsnAlaMetPheValProHisTyrAsnLeuAsnAlaSer	380

[illegible]

23	ArgGluGlnProGlnGlnAsnGluCysGlnIleGlnArgLeuAsnAlaLeuLysProGly	42
61	AATCGCATTTGAATCAGAGGCGGTATCATTTGACACTTGGAAACCCCAACACACGAGGATTC	120
43	AsnArgIleGluSerGluGlyGlyPheIleGluThrTrpAsnProAsnAsnLysProPhe	62
121	GAATGGCGCGGCTCGCCCTCTCTCGTTAGTCTCTCGCGCGAAACGCCCTTCGTAGGCT	180
63	GlnCysAlaGlyValAlaLeuSerArgCysThrLeuAsnArgAsnAlaLeuArgArgPro	82
181	TTCTACTCCAATGCTCCCAAGGAGATCTTCATCCAGCAAGGAAGGGGATACTTTGGTGTG	240
83	SerTyrThrAsnAlaProGlnGluIleTyrIleGlnGlnGlySerGlyIlePheGlyMet	102
241	ATATTCCTCGTGTCTCTAGACACTATGAAGAGCCTCACACAAGGTCGTCTGACTCTCAG	300
103	IlePheProGlyCysProSerThrPheGluGluProGlnGlnLysGly	119
301	TCCCAAGACCACCAAGACGTCTCCAAGGAGAGAACCAAGCCACACGCAACAGATAGT	360
120	SerSerArgPro	126
361	CACCAGAAGTCGACCGTTTCGATGAGGGTGATCTCATTCAGCTTCCCACCGGTGTGCT	420
127	HisGlnLysIleTyrHisPheArgGluGlyAspLeuIleAlaValProThrGlyPheAla	146
421	TTCTGGCTCTACAACGACCAACGACACTGATGTTGTGCTGTTCTCTTACTGACACCAAC	480
147	TyrTrpMetTyrAsnAsnGluAspThrLeuValValAlaValSerLeuIleAspThrAsn	166
481	AACAACGACCAACGACTTGATCAGTTCCTCCGAGAGATTCAATTTGGCTGGGAACACGGAG	540
167	SerPheGlnAsnGlnLeuAspGlnMetProArgPheTyrLeuAlaGlyAsnGlnGlu	186
541	CAAGAGTCTTAAGGTACCAACCAACAGCAACAGACAGACAGCAAGAACCTTACCATAT	600
187	GlnGluPheLeuGlnTyrGlnProGlnLysGlnGlnGly	199
601	AGCCCATACAGCCCGCAAGTCAGCCTAGACAAAGAGCGCTGAATTTAGCCCTCGAGGA	660
200	-----Gly	200
661	CAGCACAGCCGACAGAAACGAGCAGGACAAAGAGAAGAAACGAAGGTGAAACATCTTC	720
201	ThrGlnSerGlnLysGlyLysArgGlnGlnGluGluAsnGluGlyGlySerMetLeu	220
721	AGCGGCTTACGCGCGGTTCTCTGGACACAGCCCTCCAGGTTCCAGCACACAGATAGTC	780
221	SerGlyPheAlaProGluPheLeuGluHisAlaPheValVal---AspArgGlnIleVal	239
781	CAAAACCTAAGAGCGCAGACCGAGAGTGAAGAGAGGGAGCCATTGTGACAGTGAAGGGA	840
240	ArgLysLeuGlnGlyGluAsnGluGluGluLysGlyAlaIleValThrValLysGly	259
841	GGCCTCAGAACTTTGAGCCCA-----GATAGAAAGAGACGTGCCGACAGGAAGAG	891
260	GlyLeuSerValIleSerProThrGluGluGlnGlnArgProGluGluGluGlu	279
892	GAATACGATCAAGATGAAATATGATACGATGAGAGGATAGAGGCGTCCGCGGGGAGC	951
280	LysProAsp-----CysAspGluLysAspLysHisCysGlnSerGlnSer	294
952	AGAGGACGGGGGAATGGTATTGAAGAGACGATCTGCACCGCAAGTGTCTAAAAAGAACATT	1011
295	Arg-----AsnGlyIleAspGluThrIleCysThrMetArgLeuArgHisAsnIle	311
1012	GGTAGAAACAGATCCCTGCATCTCAACCCCTCAAGCTGGTTCTCACTCAAACTGCCAAC	1071
312	GlyGlnThrSerSerProAspIlePheAsnProGlnAlaGlySerIleThrThrAlaThr	331
1072	GATCTCAACCTTCAATACTAGTGGCTTGGACCTAGTCTGAATATGAAATCTCTAC	1131
332	SerLeuAspPheProAlaLeuSerTrpLeuLysLeuSerIleGlnPheGlySerLeuArg	351

Qy 1 CGGCAGCAACCGAGGAGAACCGCGTGCAGTTCAGCGCCTCAATGCGCAGACCTGAC 60

QY	1132	AGGAATGCTATGTTTGTGCTCACTACAAACACCAACGACACAGCATCATATATCGATTG	1191
Db	352	LysAsnAlaMetPheValProHisTyrAsnLeuAsnAlaAsnSerIleIleTyrAlaLeu	371
QY	1192	AGGGACGGCTCAGTCGAAGTCGTGGACAGCAACGGCAACAGAGTGTACGACGAGGAG	1251
Db	372	AsnGlyArgAlaLeuValGlnValValAsnCysAsnGlyGluArgValPheAspGlyGlu	391
QY	1252	CTTCAAGAGGTCACGTGCTGTGGTCCACAGAACTTCGCCCTCGCTGGAAAGTCCAG	1311
Db	392	LeuGlnGlyGlnValLeuIleValProGlnAsnPheAlaValAlaAlaArgSerGln	411
QY	1312	AGCGAGAACTTCGAATACGTGGCACTTCAAGACAGACTCAAGGCCAGCATAGCAACCTC	1371
Db	412	SerAspAsnPheGluTyrValSerPheLysThrAsnAspArgProSerIleGlyAsnLeu	431
QY	1372	GCGGTGAAACTCCGTCATAGATTAACCTCCGAGGAGGTGTTGCAATTCATATGGC	1431
Db	432	AlaGlyAlaAsnSerLeuLeuAsnAlaLeuProGluGluValIleGlnGlnThrPheAsn	451
QY	1432	CTCCAAGGAGGACGAGGCAAGCTTAAAGAACAAACACCCCTTCAAGTTCCTTCGTCCA	1491
Db	452	LeuArgGinglnAlaArgGlnValLysAsnAsnAsnProPheSerPheLeuValPro	471
QY	1492	CCGTCTCAGCAGTCTCCGAGGCTGTGGCT	1521
Db	472	ProLysGluSerGlnArgArgValValAla	481
RESULT 14			
ID	Q852U4	PRELIMINARY; PRT; 482 AA.	
AC	Q852U4;		
DT	01-JUN-2003 (TrEMBLrel. 24, Created)		
DT	01-JUN-2003 (TrEMBLrel. 24, Last sequence update)		
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)		
DE	Glycinin A1B2-784.		
OS	Glycine max (Soybean).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;		
OC	eurosid 1; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.		
OX	NCBI_TaxID=3847;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	STRAIN=Matsuura; TISSUE=Seed;		
RC	Fukazawa C.;		
RL	Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AB030495; BAC55938.1; -.		
DR	HSSP; P04776; 1FXZ.		
DR	GO; GO:0045735; F:nutrient reservoir activity; IEA.		
DR	InterPro; IPR006045; Cupin.		
DR	InterPro; IPR007113; Cupin region.		
DR	InterPro; IPR011051; RmlC like cupin.		
DR	InterPro; IPR006044; Seedstore_11s.		
DR	Pfam; PF00150; Cupin; 2.		
DR	PRINTS; PR00439; 1ISGL0BULIN.		
DR	PROSITE; PS00305; 11S SEED STORAGE; UNKNOWN 1.		
SQ	SEQUENCE 482 AA; 54298 MW; 79086863D946BC0D CRC64;		
Alignment Scores:			
Pred. No.:	6,49e-96	Length:	482
Score:	1488.50	Matches:	292
Percent Similarity:	71.37%	Conservative:	72
Best Local Similarity:	57.25%	Mismatches:	93
Query Match:	55.31%	Indels:	53
DB:	2	Gaps:	6
US-10-728-051-3 (1-1524) x Q852U4 (1-482)			
QY	1	CGCAGCAACCGGAGGAGACGGTGCAGTTCAGCGCCTCAATCGCGAGACCTGAC	60
Db	23	ArgGluGlnProGlnGlnAsnGluCysGlnIleGlnArgLeuAsnAlaLeuLysProGly	42

QY	61	AATCGCAATGAATCAGAGGCGGTTACATTGAGACTTGGAAACCCCAACACGAGGATTC	120
Db	43	AsnArgIleGluSerGluGlyGlyPheIleGluThrTyrAsnProAsnAsnLysProPhe	62
QY	121	GAAATGCCCGGCTGCGCTCTCTCGCTAGTCTCCCGCGCAACGCCCTTGTAGGCT	180
Db	63	GlnCysAlaGlyValAlaLeuSerArgTyrThrLeuIleArgAsnAlaLeuArgPro	82
QY	181	TTCTACTCCAAATGCTCCCGAGGAGATCTTCATCCAGCAAGGAGGATCTTTGGTTG	240
Db	83	SerTyrThrAsnAlaProGlnGluIleIleGlnGlnGlyAsnGlyIlePheGlyMet	102
QY	241	ATATTCCCTGGTGTCTAGACACTATGAAGAGCCTCACACAAAGGTGCTCGATCTAG	300
Db	103	IlePheProGlyCysProSerThrPheGluGluProGlnGlnLysGly-----Gln	119
QY	301	TTCCAAAGACCAACAGAGCTCTCCAGGAGAAGCAACCAAGCAACAGCAAGATAGT	360
Db	120	SerSerArgPro-----GlnAspArg	126
QY	361	CACCAAGAGGTGCACCGTTTCGATGAGGTGATCTCATTTGCAGTTCCCAACCGGTGTGT	420
Db	127	HisGlnLysIleTyrHisPheArgGluGlyAspLeuIleAlaValProThrGlyLeuAla	146
QY	421	TTCTGCTCTCAACAGCACCAACGACACTGATGTGTGTGTGTCTTCTTACTACACCAAC	480
Db	147	TyrTyrMetTyrAsnAsnGluAspThrLeuValValAlaValSerIleIleAspThrAsn	166
QY	481	AACAAGCAACACGAGTTCAGTTCCTCCAGGAGATTCATTTGGTGGGAACACGGAG	540
Db	167	SerLeuGluAsnGlnLeuAspGlnMetProArgArgPheTyrLeuAlaGlyAsnGlnGln	186
QY	541	CAAGAGTCTTAAGTACCAGCAACAAAGACCAAAAGCAGACGAAAGAGCTTACCATAT	600
Db	187	GlnGluPheLeuGlnTyrGlnSerGlnLysGlnGlnGly-----Gly	200
QY	601	AGCCCATACAGCCCGAAAGTCAGCTAGCAAGAGAGCGTGAATTTAGCCCTCGAGA	660
Db	200	-----Gly	200
QY	661	CAGCAGCCGAGAGACGACGACGAGCAAGAGAGAAACGAAAGTGGAAACATCTTC	720
Db	201	ThrGlnSerGlnLysGlyLysArgGlnGlnGluGluAsnGluGlyGlySerMetLeu	220
QY	721	AGCGCTTCACGCGGAGTTCCTGGAAACAAGCTTCCAGTTGACGACACAGATAGTG	780
Db	221	SerGlyPheAlaProGluPheLeuGluHisAlaPheValValAspArgGlnIleValVal	240
QY	781	CAAAACCTTAAGGCGAGACCGAGAGTGAAGAGGAGGCCATTTGTACAGTGGGGA	840
Db	241	ArgLysLeuGlnGlyGluAsnGluGluGlyLysGlyAlaIleValThrValArgGly	260
QY	841	GCGCTCAGAACTTCGAGCCCA-----GATAGAAGAGACGTGCCACCAAGAGAG	891
Db	261	GlyLeuSerValIleSerProThrGluGluArgArgGlnArgProGluGluGlu	280
QY	892	GAATACGATGAAGATCAATATCAATACGATGAAGAGGATAGAGCGTGGCAGGGAAGC	951
Db	281	LysProAsp-----CysAspGluLysAspLysHisCysGlnSerGlnSer	295
QY	952	AGAGCAGGGGAATGGTATTGAAGAGAGCATCTCGACCGCAAGTGTCTAAAGAACATT	1011
Db	296	Arg-----AsnGlyIleAspGluThrIleCysThrMetGlyLeuArgHisAsnIle	312
QY	1012	GGTAGAAACAGATCCCTGACATCTACAACCTCAAGCTGGTTCATCTCAAACTCCAC	1071
Db	313	GlyGlnThrSerSerProAspIlePheAsnProGlnAlaGlySerIleThrThrAlaThr	332
QY	1072	GATCTCAACTCTTAATACTTAGTGGCTTGGACCTAGTGTGCTGAATATGGAATCTCTAC	1131
Db	333	SerLeuAspPheProAlaLeuSerTrpLeuLysLeuSerAlaGlnPheGlySerLeuArg	352
QY	1132	AGGAATGATTTGTTGCTCACTACACCAACCAACGACGACATCATATCGATTG	1191

```

Db 353 LyAsAlaMetPheValProHisTyrAsnLeuAsnAlaAsnSerIleIleTyrAlaLeu 372
QY 1192 AGGGAGCGGCTCACGTGCAAGTCGTGAGCAGCAACGCGCAACAGAGTGTACGACGAGGAG 1251
Db 373 AsnGlyArgAlaLeuValGlnValValAsnCysAsnGlyGluArgValPheAspGlyGlu 392
QY 1252 CTTCAAGAGGGTCACGTGCTGTGGTGGCCACAGAACTTCGCCCTCGCTGGAAAGTCCAG 1311
Db 393 LeuGlnGluGlyGlnValLeuThrValProGlnAsnPheAlaValAlaAlaArgSerGln 412
QY 1312 AGCGAGAACTCCGAATACGTGGCATTCAAGACAGACTCAAGGCCCGCAGCATAGCAACCTC 1371
Db 413 SerAspAsnPheGluTyrValSerPheLysThrAsnAspArgProSerIleGlyAsnLeu 432
QY 1372 GCCGTGAAACTCCGTCTAGATTAACCTCCCGAGGAGGTGGTGCATAATTCATATGGC 1431
Db 433 AlaGlyAlaAsnSerLeuLeuAsnAlaLeuProGluGluValIleGlnGlnThrPheAsn 452
QY 1432 CTCGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1491
Db 453 LeuArgArgGlnGlnAlaArgGlnValLysAsnAsnAsnProPheSerPheLeuValPro 472
QY 1492 CCGTCTCAGAGTCTCCGAGGGCTGTGGCT 1521
Db 473 ProLysGluSerGlnArgArgValAla 482

RESULT 15
Q41702
ID Q41702 PRELIMINARY; PRT; 498 AA.
AC Q41702;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Legumin A precursor.
OS Vicia sativa (Spring vetch) (Tare).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Vicia.
OX NCBI_TaxID=3908;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cotyledon;
RA Nong V., Becker C., Muentz K.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: Hexamer; each subunit is composed of an acidic and a
CC basic chain derived from a single precursor and linked by a
CC disulfide bond (By similarity).
CC -1- SIMILARITY: Belongs to the 11S seed storage protein (globulins)
CC family.
CC EMBL; Z32835; CAA83677.1; -.
DR PIR; S44294; S44294.
DR HSP; P04776; 1FXZ.
DR GO; GO:0045735; F:nutrient reservoir activity; IEA.
DR InterPro; IPR006045; Cupin.
DR InterPro; IPR007113; Cupin region.
DR InterPro; IPR011051; RnLc like cupin.
DR InterPro; IPR006044; seedstore_11s.
DR Pfam; PF00190; Cupin; 2.
DR PRINTS; PRO0439; 11SGLOBULIN.
DR PROSITE; PS00305; 11S SEED STORAGE; 1.
KW Seed storage protein; Signal; Storage protein.
FT SIGNAL 1 21
FT CHAIN 22 498 legumin A; 11S globulin.
SQ SEQUENCE 498 AA; 56564 MW; 3575D87DB9A99699 CRC64;

Alignment Scores:
Pred. No.: 9,77e-94 Length: 498
Score: 1457.50 Matches: 289
Percent Similarity: 68.88% Conservative: 74
Best Local Similarity: 54.84% Mismatches: 93
Query Match: 54.16% Indels: 71
DB: 2 Gaps: 8

```

```

US-10-728-051-3 (1-1524) x Q41702 (1-498)
QY 1 CGGCAGCAACCGGAGAGAACCGCTGCCAGTTCCAGTTCCTCAATCGCAGAGACCTGAC 60
Db 23 ArgGluGlnSerGlnAsnGluCysGlnLeuGluArgIleAsnAlaLeuGluProAsp 42
QY 61 AATCGCATTAATCAGAGGCGGCTTACATTTAGACTTGGAAACCCCAACACGAGGAGTTC 120
Db 43 AsnArgIleGluSerGluGlyLeuIleGluThrTrpAsnProAsnAsnArgPhe 62
QY 121 GAATCGCGCGGCTGCCCTCTCTCGCTTAGTCTCCGCGCAACCCCTTCGTAGGCCT 180
Db 63 ArgCysAlaArgValAlaLeuSerArgAlaThrLeuGlnArgAsnAlaLeuArgArgPro 82
QY 181 TTCTACTCCAATGCTCCCGAGGAGATCTTCATCCAGCAAGGAGGATCTTTGGGTG 240
Db 83 TyrTyrSerAsnAlaProGlnGluIleTyrIleGlnGlnGlyAsnGlyTyrPheGlyMet 102
QY 241 ATATTCCCTGGTTCCTAGACACTATGAAGAGCTTCACACACAGGTCTCGATCTCAG 300
Db 103 ValPheProGlyCysProGluThrHisGluGluPro-----GlnGln 116
QY 301 TCCCAAGACCAACCAAGAGCTCTCCAAGGAGAGACCAAGCAACAGCAACGAGATAGT 360
Db 117 SerGlu-----GlnGlyGlu-----GlyArgArgTyrArgAspSer 128
QY 361 CACGAGAGGTGCACCGTTTCGATGAGGTGATCTCATTCGAGTTCCTCCACCGGTGTGCT 420
Db 129 HisGlnLysValAsnArgPheArgGluGlyAspIleIleAlaValProThrGlyIleAla 148
QY 421 TTCTGCTCTACAAACGACACGACACTGATGTGTGTGCTGTCTTCTTACTGACACCAAC 480
Db 149 PheTrpMetTyrAsnAspGlnAspThrProValIleAlaIleSerLeuThrAspThrGly 168
QY 481 AACACGACCAACCGCTTCATGATTCCTCCAGGAGATTCATTTGCTGGGAACACGAG 540
Db 169 SerSerAsnAsnGlnLeuAspGlnMetProArgArgPheTyrLeuAlaGlyAsnGlnGlu 188
QY 541 CAAGAGTCTTTAAGGTACCAGCAACCAAGCAGACGACAAAGCAGACGAAAGAGCTTACCATAT 600
Db 189 GlnGluPheLeuArgTyrGlnHisGln----- 197
QY 601 AGCCCATACAGCCCGCAAGTAGCTAGCCTAGACAAGAGAGCGGTGAATTTAGCCCTCGAGGA 660
Db 197 ----- 197
QY 661 CAGCAGCGCGCAGAGAACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 720
Db 198 -----GlnGlyLysGlnGlnGlnAspAsnAspGlyAsnAsnIlePhe 212
QY 721 AGCGGCTTCACGCGCGGAGTTCTTGGAAACCAAGCCTTCAGGTTTCACGACGACGACGATAGTG 780
Db 213 SerGlyPheLysArgAspPheLeuGluAspAlaPheAsnVal---AsnArgHisIleVal 231
QY 781 CAAAACCTAAGAGCGAGACCGGAGAGTGAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840
Db 232 AspArgLeuGlnGlyArgAsnGluAspGluGluLysGlyAlaIleValLysValLysGly 251
QY 841 GGCCTCAGATCTTGAGCCCGACATAGAAAG-----AGACGT 876
Db 252 GlyLeuSerIleIleAlaProGluArgGlnAlaArgHisGluArgGlySerArgGln 271
QY 877 GCGCAGCAAGAGAGGAATACGATGAA-----GAT 906
Db 272 GluGluAspGluAspGluLysGluGluArgGlnProSerHisIlySerArgArgAsp 291
QY 907 GAATATCAATACATCAAGAGGATAGA-----AGCGTGGCAGGGGAGCAACGAGGAGGAG 960
Db 292 GluAspGluAspAspLysGluLysArgHisSerGlnLysGlyGlnSerArgArgGlnGly 311
QY 961 GGAATGGTATTGAAGAGAGCATCTGCACCGCAAGTCTCAAAAAGACATTTGTTAGAAC 1020
Db ----- 1020

```

```

Db      312 AspAsnGlyLeuGluThrValCysThrAlaLysLeuArgAlaAsnIleGlySerSer 331
QY      1021 AGATCCCTGACATCTACACCTCAAGCTGGTTCACCTCAAACTGCCAACGATCTCAAC 1080
Db      332 ProSerProAspIleTyrAsnProGlnAlaGlyArgIleLysThrValThrSerLeuAsp 351
QY      1081 CTTCTAATACTTAGTGGCTGGACCTAGTGTGAATATGGAATCTCTACAGGAATGCA 1140
Db      352 LeuProValLeuArgTrpLeuLysLeuSerAlaGluHisGlySerLeuHisLysAsnAla 371
QY      1141 TTCTTTGTCTCCTCACACCAACGACACAGCATCATATATCGATTGAGGGACGG 1200
Db      372 MetPheValProHisTyrAsnLeuAsnAlaAsnSerValIleTyrAlaLeuLysGlyArg 391
QY      1201 GCTCAGCTGCAAGCTGGACAGCAACGACAGAGTGTACGACGAGGAGCTTCAAGAG 1260
Db      392 AlaArgLeuGlnValValAsnCysAsnGlyAsnThrValPheAspGlyGluLeuGluAla 411
QY      1261 GGTACAGTGTCTGTGGTCACAGAACTTCGCCGTGGTGGAAAGTCCGAGAGCGAGAAC 1320
Db      412 GlyArgAlaLeuThrValProGlnAsnTyrAlaValAlaLysSerLeuSerGluArg 431
QY      1321 TTCGAATACGTGGCATTCAAGCAGACTCAAGCCCGACGATAGCCCACTCGCCGGTGAA 1380
Db      432 PheThrTyrValAlaPheLysThrAspArgAlaSerIleAlaArgLeuAlaGlyThr 451
QY      1381 AACTCGTTCATAGTAACCTGCCGAGGAGGTGGTTGCCAAATTCATATGCGCTCCAAAGG 1440
Db      452 SerSerValIleAspAspLeuProLeuAspValValAlaAlaThrPheAsnMetGlnArg 471
QY      1441 GACGAGGCAAGGAGCTTAAGAACAAACACCCCTTCAAGTTCTTCGTTCCACCGTCTCAG 1500
Db      472 AsnGluAlaArgGlnLeuLysSerAsnAsnProPheLysPheLeuValProArgGln 491
QY      1501 CACTCTCCGAGGCTGTGGCT 1521
Db      492 SerGluMetArgAlaSerAla 498

```

Search completed: April 8, 2005, 05:25:50
 Job time : 296.329 secs